A Unifying Theory of Tests of Rank*

Majid M. Al-Sadoon

Universitat Pompeu Fabra & Barcelona GSE

April 1, 2017

Abstract

The general principles underlying tests of matrix rank are investigated. It is demonstrated that statistics for such tests can be seen as implicit functions of null space estimators. In turn, the asymptotic behaviour of the null space estimators is shown to determine the asymptotic behaviour of the statistics through a plug–in principle. The theory simplifies the asymptotics under a variety of alternatives of empirical relevance as well as misspecification, clarifies the relationships between the various existing tests, makes use of important results in the numerical analysis literature, and motivates numerous new tests. A brief Monte Carlo study illustrates the results.

JEL Classification: C12, C13, C30.

*Earlier versions of this paper were titled, "A General Theory of Rank Testing" and "The Stochastic Wald Test." I am grateful to the editor, the associated editor, and three anonymous referees for the many comments, suggestions, and corrections that have helped improved the manuscript. Thanks are also due to Sean Holly, M. Hashem Pesaran, Richard J. Smith, George Kapetanios, Rod McCrorie, Frank Kleibergen, Jesús Gonzalo, Vasco Carvalho, Kristoffer Nimark, Gábor Lugosi, Sergi Jiménez, Christian Brownlees, Geert Mesters, Barbara Rossi, Peter Boswijk, Herman van Dijk, Andrew Harvey, Luis Martins, Søren Johansen, Natercia Fortuna, Vladas Pipiras, Alan Izenman, Emre Aylar, Hanno Reuvers, Jean-Marie Dufour, Sara Nyman, and to seminar participants at Universitat Pompeu Fabra, Universidad Carlos III de Madrid, Universidade do Minho, Universidade do Porto, and Lund University for helpful comments and suggestions. The author takes sole responsibility for any remaining errors. Some of the results of this paper were part of the author's PhD thesis. This paper is dedicated to T. W. Anderson (1918–2016), who made the seminal contribution to this field. Research for this paper was supported by Spanish Ministry of Economy and Competitiveness projects ECO2012-33247 and ECO2015-68136-P (MINECO/FEDER, UE) and Fundación BBVA Scientific Research Grant PR16-DAT-0043.

Keywords: Tests of rank, cointegration, plug–in principle, null space estimation, reduced–rank approximation, local power, misspecification.

1 Introduction

The literature on tests of matrix rank has grown into a large and eminently applicable branch of econometrics since the seminal contribution by Anderson (1951) (see Camba-Mendez & Kapetanios (2009) for a survey). Much of this progress has taken place in spite of the difficulty of the asymptotics of these tests; indeed, statistics for testing the rank of a matrix often involve eigenvectors, inverses, and other discontinuous functions of matrices. Consequently, significant gaps have persisted in our knowledge of the relationships between the various tests in the literature. Relationships between the Anderson (1951), Johansen (1991), Robin & Smith (2000), and Kleibergen & Paap (2006) statistics are known, as are relationships between the Cragg & Donald (1996, 1997) statistics. However, there is as of yet no characterization of the general structure of statistics for testing the rank of a matrix. Another consequence, is that little is known about the behaviour of tests of rank under local alternatives or under misspecification. Local power is considered in Cragg & Donald (1997) and a handful of papers surveyed by Hubrich et al. (2001), while misspecification is considered in Robin & Smith (2000), Caner (1998), Cavaliere et al. (2010b), Aznar & Salvador (2002), and Cavaliere et al. (2014). All of these results relate to specific tests and there is as of yet no known general principle that unifies all of these results. The statistical and econometric literature has also made little use of the numerical analysis literature, which has made great strides in understanding and discovering effective matrix rank (Hansen, 1998).

Thus, the objective of this paper is to investigate the general principles underlying tests of matrix rank by: (i) characterizing the general structure of statistics for testing matrix rank, (ii) describing the behaviour of these statistics under a variety of alternatives of empirical relevance and misspecification, and (iii) making use of important insights from the numerical analysis literature. These intermediate objectives are achieved along the following steps.

First, the paper shows that the general structure of statistics for tests of rank is of the form of an implicit function of estimators of the null spaces of the matrix in question (see Sections 3.1 and 4.3). This achieves intermediate objective (i) as it is demonstrated that most statistics in the literature have common functional forms although they may differ in the implicit null space estimators (see Table 1).

Next, the paper develops the theory of null space estimation based on reduced–rank approximations, which have been widely studied in the numerical analysis literature. This achieves intermediate objective (iii). Lemmas 1 and 2 provide a full characterization of the asymptotic properties of null space estimators under the various alternatives under study. These results generalize Dufour & Valery (2011) in that they apply to general matrices rather than just the positive semi–definite ones and are not restricted to eigenprojections. They also allow us to use any reduced–rank approximation to construct a statistical test of matrix rank; this is demonstrated by a number of new tests based on the QR and Cholesky decompositions (see Section 5).

Finally, it is demonstrated that the behaviour of statistics for tests of rank is completely governed by the implicit null space estimators. A plug–in principle is shown to hold, whereby every statistic mimics the asymptotic behaviour of an infeasible statistic that plugs in null spaces related to the population value of the matrix under study. This greatly simplifies the asymptotics of tests of rank under the various alternatives as well as misspecification. Under the null hypothesis or the local alternative, one can simply ignore the fact that the null spaces are estimated and derive the asymptotics as if the appropriate null spaces were known. Under the global alternative, the statistic diverges whenever the associated infeasible statistic diverges and under certain conditions (conjectured to be generic) both statistics are proven to diverge at the same rate. Thus, the plug–in principle allows us to achieve intermediate objective (ii). It also follows that statistics that have a common functional form but differ in their null space estimators are asymptoticallly equivalent, therefore establishing the asymptotic equivalence of a number of tests in the literature. Theorem 2 and Corollaries 3 and 4 are shown to imply the asymptotics of almost all tests of rank, with the handful of exceptions demonstrably satisfying a weaker form of the plug–in principle (see the discussion in Section 4.3).

It is important to emphasize several distinctive features of the approach of this paper. First, the approach is Waldian in that the primitives are taken to be a matrix estimator and a normalizing matrix; this allows it to encompass a much wider variety of tests than Reinsel & Velu (1998) and Massmann (2007), which nest some of the likelihood–based tests but miss a host of other tests. Second, it is based on orthogonal projection matrices, so that no identifying restrictions are imposed on the null space estimators; this allows for an elegant and compact description of their rates of convergence. Third, it encompasses both standard asymptotics and cointegration in a way that illuminates the continuity between the two literatures. In this regard, the paper is developed gradually from the special case of standard asymptotics to the general case that allows for cointegration.

It is also important to note two aspects of the plug–in principle that have been well known in the literature. First, as far back as Stock & Watson (1988) and as recently as Boswijk et al. (2015), researchers have relied on the idea that the population cointegration relationship could be substituting in for a super–consistent estimator in working out the asymptotics of cointegration statistics. This paper demonstrates that this idea does not hold in general (see Example 4) and proposes the necessary modifications. Second, the proofs of the asymptotics of some tests sometimes involved an implicit use of the plug–in principle (e.g. Cragg & Donald (1996) and Robin & Smith (2000)). However, these instances concerned specific rather than generic tests and did not recognize the plug–in principle as an overarching framework that elucidates the asymptotics of tests of rank in general.

In terms of practical recommendations for practitioners, the following results emerge: (i) both theoretical and Monte Carlo results fail to point to an optimal test of rank, thus researchers can base their choice of test on other considerations, (ii) test statistics based on the QR and LU decompositions (e.g. the Cragg & Donald (1996) statistic) are recommended for high intensity computing such as the bootstrap as they are numerically less expensive than the alternatives (see Al-Sadoon (2016) for an illustration), and (iii) the paper proposes a number of new tests, which include robust extensions of the likelihood ratio test of Anderson (1951) and the maximum eigenvalue test of Johansen (1991) as well as tests based on the QR and Cholesky decompositions.[1]

The paper is organized as follows. Section 2 develops the notation of the paper. Section 3 develops the theory under standard asymptotics. Section 4 develops the theory under non–standard asymptotics. Section 5 provides Monte Carlo evidence. Section 6 concludes. Further Monte Carlo results and technical material as well as the proofs of the results can be found in the on-line appendix to the paper.

2 Notation

$\mathbb{R}^{n\times m}$ denotes the set of $n \times m$ real valued matrices and $\mathbb{G}^{n\times m}$ is the subset of matrices of full rank. $\mathbb{P}^m_+ \subset \mathbb{P}^m \subset \mathbb{S}^m$ denote the set of positive definite, positive semi–definite, and symmetric matrices in $\mathbb{R}^{m\times m}$ respectively. $\mathrm{vec}(B)$ is the vector formed by vertically stacking the columns

[1]Practitioners may also wish to consult the Matlab tutorial accompanying this paper, `tutorial.m`, which is included in the compressed file, `rank.rar`, available on the author's website.

of B and $\text{vech}(B)$ is the one formed by vertically stacking the elements below and including the diagonal elements of B. The mat operator is defined as the inverse to the vec operator (its range will be evident from the context). The Euclidean norm of $B \in \mathbb{R}^{n\times m}$ is defined as $\|B\| = (\text{vec}'(B)\text{vec}(B))^{1/2}$. The Mahalanobis norm is defined as $\|B\|_\Theta = (\text{vec}'(B)\Theta^{-1}\text{vec}(B))^{1/2}$ for $\Theta \in \mathbb{P}_+^{nm}$. The 2–norm is defined as $\|B\|_2 = \max_{x\in\mathbb{R}^m, \|x\|=1} \|Bx\|$. If $\mathcal{P} \subset \mathbb{R}^{n\times m}$, define $d(B, \mathcal{P}) = \inf_{X\in\mathcal{P}} \|B - X\|$. The singular values of B are denoted by $\sigma_1(B) \geq \sigma_2(B) \geq \cdots \geq \sigma_{\min\{n,m\}}(B)$. The condition number of B is defined as $\text{cond}(B) = \sigma_1(B)/\sigma_r(B)$, where $r = \text{rank}(B)$. The Moore–Penrose inverse of B is denoted by $B^\dagger$. For any $B \in \mathbb{G}^{n\times m}$ with $n > m$, an orthogonal complement $B_\perp$ is any matrix in $\mathbb{G}^{n\times(n-m)}$ satisfying $B'_\perp B = 0$. The column space of B is denoted by $\text{span}(B)$. The orthogonal projection onto $\text{span}(B)$ is denoted by P_B. The duplication matrix D_m is the mapping $\text{vech}(B) \mapsto \text{vec}(B)$ over $B \in \mathbb{S}^m$. For $B \in \mathbb{P}^m$, $B^{1/2}$ is the positive semi–definite square root matrix and $B^{\dagger/2} = (B^{1/2})^\dagger = (B^\dagger)^{1/2}$.

Finally, we say that a sequence of random matrices $X_T \in \mathbb{R}^{n\times m}$ indexed by T is bounded away from zero in probability and denote this by $X = O_p^{-1}(1)$ if for all $\varepsilon > 0$, there exists a $\delta_\varepsilon > 0$ and a $T_\varepsilon \geq 0$ such that the probability that $\|X_T\| > \delta_\varepsilon$ is at least $1 - \varepsilon$ for all $T \geq T_\varepsilon$. It is easy to show that $\|X_T\|^{-1} = O_p^{-1}(1)$ if and only if $X_T = O_p(1)$ and $X_T = O_p^{-1}(1)$ if and only if $\|X_T\|^{-1} = O_p(1)$. Hence the notation, $O_p^{-1}(1)$. The product of two $O_p^{-1}(1)$ sequences is again $O_p^{-1}(1)$ and $a_T\|X_T\| \xrightarrow{p} \infty$ for any non–random sequence $a_T \to \infty$. The deterministic version, $O^{-1}(1)$, is defined similarly.

3 Tests of Rank Under Standard Asymptotics

This section lays the foundations of our study. First, the general structure of statistics for tests of rank is considered. It is shown that most of them are implicit functions of null space estimators. Second, the behaviour of null space estimators is investigated under standard asymptotic assumptions. Finally, it is shown how their behaviour governs the asymptotics of tests of rank. Before we begin, we must fix a few ideas.

We will draw inference on an unobserved matrix B in a parameter space that will be taken to be either $\mathbb{R}^{n\times m}$, $\mathbb{S}^m$, or $\mathbb{P}^m$ (the particular parameter space will be evident from the context). For $0 \leq r < \min\{n, m\}$, we will be interested in testing the hypothesis

$$H_0(r) : \text{rank}(B) = r$$

against the global alternative

$$H_1(r) : \text{rank}(B) > r$$

as well as the local alternative

$$H_T(r) : B = B^* + D/\sqrt{T}, \quad \mathrm{rank}(B^*) = r,$$

where B^* is unknown, T is the sample size, and the choice of D ensures that B remains in the parameter space. Because our results depend crucially on B^* under $H_T(r)$ and on B under $H_0(r)$ and $H_1(r)$, we define $B^* = B$ under $H_0(r)$ and $H_1(r)$. Thus, the discussion below will focus primarily on B^*, which is to be understood as B under $H_0(r)$ and $H_1(r)$ and the limiting B under $H_T(r)$.

3.1 Preliminary Examples

First, we review the core statistics in the literature that form the basis of our investigation.

Example 1. Let $\{(y_t, x_t) : t \geq 1\}$ be stationary and ergodic with finite fourth moments and

$$y_t = Bx_t + \varepsilon_t, \qquad t = 1, \dots, T,$$

where B is an element of the parameter space $\mathbb{R}^{n\times m}$. Assume, moreover, that $\Gamma = E(x_t x_t') \in \mathbb{P}_+^m$ and $\{x_t \otimes \varepsilon_t : t \geq 0\}$ is a martingale difference sequence such that $E(x_t x_t' \otimes \varepsilon_t \varepsilon_t') \in \mathbb{P}_+^{nm}$. Let $\widehat{B}$ and $\widehat{\Sigma}$ be the OLS estimators of B and $E(\varepsilon_t\varepsilon_t')$ and recall that the homoskedastic estimator of the asymptotic variance of $\widehat{B}$ is $\widehat{\Omega} = \widehat{\Gamma}^{-1} \otimes \widehat{\Sigma}$, where $\widehat{\Gamma} = T^{-1}\sum_{t=1}^T x_t x_t'$. The likelihood ratio statistic proposed by Anderson (1951) for testing $H_0(r)$ can be expressed as

$$LR\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) = T \sum_{i=1}^{\min\{n,m\}} \log\left(1 + \sigma_i^2\left((P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2} P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right)\right),$$

where $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ estimate, in a sense to be made precise in the next subsection, the null spaces of B under $H_0(r)$. Anderson noted that, under $H_0(r)$, LR behaves identically to a quadratic form that is expressible as

$$A\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) = T\left\|(P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2} P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right\|^2.$$

We will also consider the Johansen (1991) maximum eigenvalue statistic, expressible as

$$J\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) = T\log\left(1 + \left\|(P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2} P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right\|_2^2\right).$$

Note that each of these statistics is a function of $\mathrm{vec}\left((P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2} P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right) = \left((P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\right)^{\dagger/2} \mathrm{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r})$. Thus, if we wish to utilize an alternative estimator to $\widehat{\Omega}$ that is not necessarily of Kronecker product form (e.g. the heteroskedasticity–robust estimator), we may utilize the following generalization of the A statistic,

$$F\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) = T\mathrm{vec}'(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r})\{(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\}^{\dagger}\mathrm{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}).$$

The statistics of Cragg & Donald (1996) and Kleibergen & Paap (2006) (among many others) are of this form. Note that $A\left(\widehat{B},\widehat{\Sigma},\widehat{\Gamma},P_{\widehat{N}_r},P_{\widehat{M_r}}\right) = F\left(\widehat{B},\widehat{\Gamma}^{-1}\otimes\widehat{\Sigma},P_{\widehat{N}_r},P_{\widehat{M_r}}\right)$. On-line appendix E.1 provides detailed derivations of all of the statistics above.

Following the same logic that leads to the F statistic, we may generalize LR and J as

$$\begin{aligned}
&LR_g\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M_r}}\right)\\
&\quad= T\sum_{i=1}^{\min\{n,m\}}\log\left(1+\sigma_i^2\left(\operatorname{mat}\left(((P_{\widehat{M_r}}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M_r}}\otimes P_{\widehat{N}_r}))^{\dagger/2}\operatorname{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M_r}})\right)\right)\right)\\
&J_g\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M_r}}\right) = T\log\left(1+\left\|\operatorname{mat}\left(((P_{\widehat{M_r}}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M_r}}\otimes P_{\widehat{N}_r}))^{\dagger/2}\operatorname{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M_r}})\right)\right\|_2^2\right).
\end{aligned}$$

These statistics reduce to the LR and J statistics when $\widehat{\Omega}=\widehat{\Gamma}^{-1}\otimes\widehat{\Sigma}$. □

Example 2. Suppose the parameter space in Example 1 is $\mathbb{S}^m$ instead. Suppose, moreover, that B is estimated subject to the symmetry restriction so that $\widehat{B}\in\mathbb{S}^m$ and $\widehat{\Psi}$ is an estimator of the asymptotic covariance of $\operatorname{vech}(\widehat{B})$. Then Donald et al. (2007) formulate a test of $H_0(r)$ using a statistic expressible as $F\left(\widehat{B},D_m\widehat{\Psi}D_m',P_{\widehat{M_r}},P_{\widehat{M_r}}\right)$, where $\widehat{M_r}$ again estimates the null space of B under $H_0(r)$. When the parameter space is $\mathbb{P}^m$, Donald et al. (2007) propose a statistic based on the eigenvalues of $\widehat{B}$ expressible as

$$t\left(\widehat{B},\widehat{\Psi},P_{\widehat{M_r}}\right)=\frac{\sqrt{T}\operatorname{tr}(P_{\widehat{M_r}}\widehat{B}P_{\widehat{M_r}})}{\sqrt{\operatorname{vec}'(I_m)(P_{\widehat{M_r}}\otimes P_{\widehat{M_r}})D_m\widehat{\Psi}D_m'(P_{\widehat{M_r}}\otimes P_{\widehat{M_r}})\operatorname{vec}(I_m)}}.$$

Note that $t\left(\widehat{B},\frac{1}{m-r}(D_m'D_m)^{-1},P_{\widehat{M_r}}\right)=\sqrt{T}\operatorname{tr}(P_{\widehat{M_r}}\widehat{B}P_{\widehat{M_r}})$. On-line appendix E.2 provides detailed derivations of the statistics above. □

The statistics LR, A, J, F, and t comprise almost all of the statistics in the literature (see Table 1).[2] The remaining statistics await further discussion below. For now, we simply note that each statistic is of the form $T^\theta\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M_r}})$, for some function τ and associated scaling factor $\theta>0$ (see Assumptions K). In fact, each is of the form of a standardized measure of the size of $P_{\widehat{N}_r}\widehat{B}P_{\widehat{M_r}}$ ($P_{\widehat{M_r}}\widehat{B}P_{\widehat{M_r}}$ in the symmetric case). Small values of this measure lend support to $H_0(r)$, while large values indicate that $H_1(r)$ is true. We will see how this works in the next section. We will also see that all of these statistics satisfy a plug–in principle whereby their asymptotic behaviour under standard asymptotics mimics the behaviour of infeasible analogues $T^\theta\tau(\widehat{B},\widehat{\Omega},P_{N_r},P_{M_r})$ where N_r and M_r are determined by B.

[2]It is worth emphasizing that while the expressions above serve the purpose of highlighting the common structure of statistics for tests of rank, they are to be avoided for computational reasons. The corresponding representations that avoid projection matrices and generalized inverses can be found in on-line appendices E.1 and E.2.

3.2 Estimating the Null Spaces

In this section, our task will be to estimate the null spaces of $B^* \in \mathbb{R}^{n\times m}$. Throughout this section, we maintain the assumption that $\widehat{B} \in \mathbb{G}^{n\times m}$. This is guaranteed if $\text{vec}(\widehat{B})$ is a non–degenerate random vector (i.e. it has a continuous probability density function). It is also guaranteed if $\text{vech}(\widehat{B})$ is non–degenerate when $\widehat{B}$ is restricted to $\mathbb{S}^m$ or $\mathbb{P}^m$. This is due to the fact that rank deficient matrices occupy sub-manifolds of the parameter space of lower dimension and therefore of measure zero.

Definition 1 (Null Space Estimators)**.** For $B^* \in \mathbb{R}^{n\times m}$, a rank–$r$ left (resp. right) null space estimator of B^* is defined as any random matrix $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ (resp. $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$) almost surely. We say that $\widehat{N}_r$ (resp. $\widehat{M}_r$) is consistent for the left (resp. right) null space of B^* if $P_{\widehat{N}_r}$ (resp. $P_{\widehat{M}_r}$) converges in probability to the orthogonal projection onto the left (resp. right) null space of B^*. □

The problem of estimating null spaces has a long history in the numerical analysis literature (Golub & Van Loan, 1996; Hansen, 1998). The basic idea is to approximate $\widehat{B}$ by a reduced–rank approximation (RRA), $\widehat{B}_r^{RRA}$ of rank r. If the approximation is good enough, then as $\widehat{B}$ approaches B^*, so will $\widehat{B}_r^{RRA}$. We may then obtain consistent estimates of the null spaces of B^* as the null spaces of $\widehat{B}_r^{RRA}$.

There are essentially two types of RRAs: decomposition–based approximations and norm–based approximations. We discuss them briefly in turn. A more detailed discussion is relegated to on-line appendix B.

Definition 2 (Decomposition–based Approximations)**.** For $\widehat{B} \in \mathbb{G}^{n\times m}$, the decomposition $\widehat{B} = \widehat{U}\widehat{S}\widehat{V}'$ is rank–revealing if, $\widehat{S} \in \mathbb{R}^{n\times m}$ is upper triangular, $\widehat{U}$ and $\widehat{V}$ and their inverses are bounded, and whenever $\widehat{S}$ is partitioned as $\begin{bmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & \widehat{S}_{22} \end{bmatrix}$ with $\widehat{S}_{11} \in \mathbb{R}^{r\times r}$, then

(i) There is a $K_1 > 0$, not dependent on $\widehat{B}$, such that $\sigma_r(\widehat{S}_{11}) \geq K_1\sigma_r(\widehat{B})$.

(ii) There is a $K_2 > 0$, such that $\sigma_1(\widehat{S}_{22}) \leq K_2\sigma_{r+1}(\widehat{B})$ and $K_2 = O(1)$ for any $\widehat{B}$ satisfying $\widehat{B} - \widehat{B}^* \to 0$ with $\widehat{B}^* = O(1)$, $\text{rank}(\widehat{B}^*) = r$, and $\sigma_r(\widehat{B}^*) = O^{-1}(1)$.

A rank–revealing decomposition can be used to obtain the RRA

$$\widehat{B}_r^{DBA} = \widehat{U}\begin{bmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & 0 \end{bmatrix}\widehat{V}'.$$

We refer to this RRA as a decomposition–based approximation (DBA). □

The set of rank–revealing decompositions includes the singular value decomposition (SVD), the LU decomposition with complete pivoting (LU), and the block LU decomposition (BLU). When $\widehat{B}$ is restricted to $\mathbb{S}^m$, the spectral decomposition (EIG) is also rank–revealing. All of these decompositions have appeared in tests of rank (see Table 1). There are many more that have not been utilized in statistics or econometrics. Examples include the QR decomposition with pivoting, which is the preferred decomposition in numerical analysis due to its computational expediency, and, when $\widehat{B}$ is restricted to $\mathbb{P}^m$, the Cholesky rank–revealing decomposition. These are illustrated in the Monte Carlo simulations of Section 5. The fact that all of these decompositions are rank–revealing is demonstrated in on-line appendix B.1.[3]

Definition 3 (Norm–based Approximations)**.** For $\widehat{B} \in \mathbb{G}^{n\times m}$, let

$$\widehat{B}_r^{CDA} \in \operatorname{argmin}\{\|\widehat{B} - A\|_\Theta : A \in \mathbb{R}^{n\times m}, \operatorname{rank}(A) \leq r\},$$

where $\Theta \in \mathbb{P}_+^{nm}$. We term this the Cragg and Donald approximation (CDA), after Cragg & Donald (1997), who first proposed it in econometrics.[4] $\square$

The CDA nests a number of other RRAs as special cases. When Θ is the identity matrix we obtain the SVD approximation. When Θ is a Kronecker product of square matrices of sizes m and n, we obtain the RRA implicit in the Robin & Smith (2000) decomposition (RSD). The null space estimators in Example 1 are obtained from $\widehat{B}$ by an RSD with $\Theta = \widehat{\Gamma}^{-1} \otimes \widehat{\Sigma}$. When Θ is not of Kronecker product form, there are no known analytical solutions. However, a novel iterative scheme for obtaining the CDA can be found in on-line appendix B.2, which works quite well in numerical experiments.

To summarize, Table 1 lists the implicit null space estimators in a selection of statistics in the literature. We are now able to extract the following lemma which describes the asymptotics of null space estimators in general.

Lemma 1. Let $\widehat{B}$ be an estimator of $B^* \in \mathbb{R}^{n\times m}$ such that $\widehat{B} \in \mathbb{G}^{n\times m}$, and $\sqrt{T}(\widehat{B} - B^*) = O_p(1)$. Let $\operatorname{rank}(B^*) = r$ and let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of B^* respectively. Let the RRAs $\{\widehat{B}_i^{RRA} : 0 \leq i < \min\{n,m\}\}$ be either DBAs or CDAs. In the latter case, we assume that $\operatorname{cond}(\Theta) = O_p(1)$. Finally, let $\widehat{N}_i$ (resp. $\widehat{M_i}$) span the left (resp. right) null space of $\widehat{B}_i^{RRA}$ for $0 \leq i < \min\{n,m\}$.

[3]Of course, not every matrix decomposition is rank–revealing. Example of non–rank–revealing decompositions include the Jordan canonical form for square matrices, the LU decomposition with partial or no pivoting, and the QR decomposition with no pivoting (see on-line appendix B.1 for a discussion).

[4]A precursor to this RRA is the one proposed by Gabriel & Zamir (1979), although they take Θ to be diagonal.

Table 1: Reduced–Rank Approximations Utilized in Statistics for Tests of Rank.

Paper	Statistic	RRA	Paper	Statistic	RRA
Bartlett (1947)	F	RSD	Anderson (1951)	LR	RSD
Anderson (1951)	A	RSD	Johansen (1988)	F	RSD
Stock & Watson (1988)	NA†	EIG	Johansen (1991)	J	RSD
Kleibergen & van Dijk (1994)	F	BLU	Robin & Smith (1995)	A	BLU
Cragg & Donald (1996)	F	LU	Yang & Bewley (1996)	F	RSD
Yang & Bewley (1996)	κ‡	RSD	Cragg & Donald (1997)	F	CDA
Bierens (1997)	κ‡	RSD	Lucas (1997)	NA†	BLU
Lucas (1998)	NA†	BLU	Caner (1998)	F	RSD
Quintos (1998)	F	RSD	Gonzalo & Pitarakis (1999)	κ‡	RSD
Lütkepohl & Saikkonen (1999)	F	RSD	Robin & Smith (2000)	κ‡	RSD
Nyblom & Harvey (2000)	t	EIG	Boswijk & Lucas (2002)	NA†	BLU
Breitung (2002)	t	EIG	Kleibergen & Paap (2006)	F	SVD
Donald et al. (2007)	F	LU	Donald et al. (2007)	F	SVD
Donald et al. (2007)	F	CDA	Donald et al. (2007)	t	EIG
Avarucci & Velasco (2009)	F	SVD	Cavaliere et al. (2010a)	F	RSD
Cavaliere et al. (2010b)	F	RSD	Nielsen (2010)	t	EIG
Johansen & Nielsen (2012)	NA†	RSD	Cavaliere et al. (2014)	F	RSD

†These statistics are not of any form we have considered in this paper. They are discussed in Section 4.
‡These statistics satisfy Assumptions K.

(i) $\sqrt{T}(\widehat{B} - \widehat{B}_r^{RRA})$, $\sqrt{T}(P_{\widehat{N}_r} - P_{N_r})$, and $\sqrt{T}(P_{\widehat{M}_r} - P_{M_r})$ are $O_p(1)$.

(ii) If $0 \leq i < r$ then $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$. If $n = m$ and $B^* \in \mathbb{P}^m$, then $P_{\widehat{M}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ and $P_{\widehat{N}_i}\widehat{B}P_{\widehat{N}_i} = O_p^{-1}(1)$.

(iii) If $0 \leq i < r$ and the rank–i RRA is continuous at B^*, then $P_{\widehat{N}_i} - P_{N_i^*} = o_p(1)$ and $P_{\widehat{M}_i} - P_{M_i^*} = o_p(1)$, where N_i^* and M_i^* span the left and right null spaces of $(B^*)_i^{RRA}$ respectively.

Lemma 1 (i) establishes the rates of convergence of DBAs, CDAs, and their associated null space estimators. Since the Euclidean norm of the difference of two orthogonal projection matrices of the same rank is the Euclidean norm of the vector of sines of all of the canonical angles between the two subspaces (Stewart & Sun, 1990), Lemma 1 (i) proves that the canonical angles between the estimated and population null spaces are $O_p(T^{-1/2})$. Dufour & Valery (2011) obtain the special case of this result where $B^*, \widehat{B} \in \mathbb{P}^m$ and $\widehat{B}_i^{RRA}$ is the EIG RRA.

Lemma 1 (ii) states that when one estimates null spaces of dimensions that are too high, the null space estimators capture non–vanishing components of $\widehat{B}$ in the sense that $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}$ remains bounded away from zero in probability. When $\widehat{B}$ approaches $\mathbb{P}^m$, the left and right null space estimators are equally capable of capturing non–vanishing components of $\widehat{B}$.

Lemma 1 (iii) states that if the RRA is continuous at B^*, then the null space estimators converge in probability. It follows that the non–vanishing components of $\widehat{B}$ can be estimated consistently as $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} \xrightarrow{p} P_{N_i}B^*P_{M_i} \neq 0$. Unfortunately, RRAs are not continuous everywhere. However, continuity is known to be generic for the SVD, RSD, and EIG RRAs (Stewart & Sun, 1990; Markovsky, 2012) and it can also be shown to be generic for simple DBAs such as the LU, Cholesky, and QR RRAs. No results are available for the general CDA, although one might well conjecture that continuity is generic for all RRAs.

3.3 The Plug–in Principle

We are now in a position to see how null space estimators determine the behaviour of statistics for tests of rank. Consider first the following set of assumptions.

Assumptions A. $B^* \in \mathbb{R}^{n\times m}$. $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{S}^{nm}$ are estimators indexed by T. Each $\text{vec}(\widehat{B}) \in \mathbb{R}^{nm}$ is a non–degenerate random vector. $\widehat{\Omega} \in \mathbb{P}_+^{nm}$ almost surely. $\sqrt{T}(\widehat{B} - B^*)$, $\widehat{\Omega}$, and $\widehat{\Omega}^{-1}$ are $O_p(1)$.

Assumptions A are satisfied in the context of Example 1. They also arise in much more general settings, e.g. generalized method of moments and maximum likelihood estimation. Note that $\widehat{\Omega}$ may or may not be a consistent estimator of the asymptotic variance of $\widehat{B}$. In fact, it is not even required to converge, allowing for the fixed–b hypothesis testing framework (Kiefer et al., 2000; Kiefer & Vogelsang, 2002a,b, 2005).

We will also want to prove results for symmetric matrices, which we considered in Example 2. In that case, we will rely on the following analogous set of assumptions.

Assumptions B. $B^* \in \mathbb{S}^m$. $\widehat{B} \in \mathbb{S}^m$ and $\widehat{\Psi} \in \mathbb{S}^{m(m+1)/2}$ are estimators indexed by T. Each $\text{vech}(\widehat{B}) \in \mathbb{R}^{m(m+1)/2}$ is a non–degenerate random vector. $\widehat{\Psi} \in \mathbb{P}_+^{m(m+1)/2}$ almost surely. $\sqrt{T}(\widehat{B} - B^*)$, $\widehat{\Psi}$, and $\widehat{\Psi}^{-1}$ are $O_p(1)$. In this context, we will set $\widehat{\Omega} = D_m\widehat{\Psi}D_m'$.

The alternatives $H_0(r)$, $H_T(r)$, and $H_1(r)$ interact with Assumptions A and B by setting the rank of B^*. In particular, under either Assumptions A or B, $\text{rank}(B^*) = r$ under $H_0(r)$ or $H_T(r)$ and $\text{rank}(B^*) > r$ under $H_1(r)$.

Under these remarkably minimal assumptions, we will be able to prove the following plug–in principle for a large class of statistics

Definition 4 (The Plug–in Principle in Standard Asymptotics)**.** Suppose $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{P}^{nm}$ are estimators indexed by T, $B^* \in \mathbb{R}^{n\times m}$, and $\sqrt{T}(\widehat{B} - B^*) = O_p(1)$. For a given

$0 \leq r < \min\{n, m\}$ and RRA, let $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $\widehat{B}_r^{RRA}$ respectively. The weak plug–in principle for statistics of tests of rank is said to hold for the statistic $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$ relative to the null spaces of B^* if

(i) Under either $H_0(r)$ or $H_T(r)$, $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) - T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r}) = O_p(T^{-1/2})$, where $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of B^* respectively.

(ii) Under $H_1(r)$, then $|\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})| = O_p^{-1}(1)$ if $|\tau(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r})| = O_p^{-1}(1)$, where $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $(B^*)_r^{RRA}$ respectively.

It is said to satisfy the strong plug–in principle relative to the null spaces of B^* if additionally

(iii) Under $H_1(r)$, $\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) - \tau(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r}) = o_p(1)$, where $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $(B^*)_r^{RRA}$ respectively. $\square$

Condition (i) requires that the feasible and infeasible statistics differ from each other by no more than $O_p(T^{-1/2})$ under $H_0(r)$ and $H_T(r)$. This is much stronger than asymptotic equivalence in large sample statistics, which requires only that the two have the same limiting distribution (Lehmann & Romano, 2005, p. 577). We will see, however, that it is easily satisfied. Condition (ii) ensures that a test of rank has power against $H_1(r)$ if the associated infeasible test has power. Condition (iii) strengthens (ii) in that it requires the feasible and infeasible statistics to diverge at the same rate under the global alternative.

The variant of the plug–in principle we will prove applies to the class of statistics of the form $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) = T^\theta \kappa\left(P_{\widehat{N}_r} \widehat{B} P_{\widehat{M}_r}, (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\right)$, where κ satisfies the following set of assumptions.

Assumptions K. $\mathcal{P} \subseteq \mathcal{X} \subseteq \mathbb{R}^{n\times m}$. $\mathcal{P}$ is closed and convex. $\mathcal{Y} \subseteq \mathbb{R}^{nm\times nm}$. $\kappa : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a measurable function and satisfies:

(i) There exist measurable functions L_1 and L_2 such that for all $X, \widehat{X} \in \mathcal{X}$ and $Y, \widehat{Y} \in \mathcal{Y}$,

$$|\kappa(\widehat{X}, \widehat{Y}) - \kappa(X, Y)| \leq L_1(\widehat{X}, X, \widehat{Y}, Y)\|\widehat{X} - X\| + L_2(\widehat{X}, X, \widehat{Y}, Y)\|\widehat{Y} - Y\|.$$

For $\theta > 0$, $L_1(\widehat{X}, X, \widehat{Y}, Y) = O(\|X\|^{2\theta-1})$ and $L_2(\widehat{X}, X, \widehat{Y}, Y) = O(\|X\|^{2\theta})$ if $\|Y^\dagger\| = O(1)$ as $\|\widehat{X} - X\| + \|\widehat{Y} - Y\| \to 0$.

(ii) For every $C_1 > 0$ and $C_2 > 0$ there exists a $C > 0$ such that for all $X \in \mathcal{P}$ and $Y \in \mathcal{Y}$ with $\mathrm{vec}(X) \in \mathrm{span}(Y)$, $\|X\| \geq C_1$ and $\|Y\| \leq C_2$ imply that $|\kappa(X, Y)| \geq C$.

Table 2: Properties of the Most Common Statistics for Tests of Rank.

τ	θ	$\mathcal{P}$	$\mathcal{X}$	$\mathcal{Y}$	$\kappa(X,Y)$
LR	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^m \otimes \mathbb{P}^n$	$\sum_{i=1}^{\min\{n,m\}} \log(1+\sigma_i^2(\text{mat}(Y^{\dagger/2}\text{vec}(X))))$
LR_g	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^{nm}$	$\sum_{i=1}^{\min\{n,m\}} \log(1+\sigma_i^2(\text{mat}(Y^{\dagger/2}\text{vec}(X))))$
A	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^m \otimes \mathbb{P}^n$	$\lVert Y^{\dagger/2}\text{vec}(X)\rVert^2$
F	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^{nm}$	$\lVert Y^{\dagger/2}\text{vec}(X)\rVert^2$
J	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^m \otimes \mathbb{P}^n$	$\log\left(1+\lVert\text{mat}(Y^{\dagger/2}\text{vec}(X))\rVert_2^2\right)$
J_g	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^{nm}$	$\log\left(1+\lVert\text{mat}(Y^{\dagger/2}\text{vec}(X))\rVert_2^2\right)$
t	$\frac{1}{2}$	$\mathbb{P}^m$	$\mathbb{S}^{m\times m}$	$\mathbb{P}^{m(m+1)/2}$	$\frac{\text{tr}(X)}{\text{vec}'(I_m)Y\text{vec}(I_m)}$

The Lipschitz condition of Assumptions K (i) allows the weak plug–in principle to hold under the null and local alternatives. It also allows the strong plug–in principle to hold under the global alternative. Assumption K (i) also reveals that the scaling factor, θ, is determined by the shape of the test statistic at the origin. On the other hand, the boundedness condition in Assumption K (ii) allows tests based on the feasible and infeasible statistics to have power against $H_1(r)$. Note that boundedness of κ away from zero is only ensured on $\mathcal{P}\times\mathcal{Y}$ rather than the potentially larger set $\mathcal{X}\times\mathcal{Y}$. This is to allow for the t test of Example 2, which has power against positive semi–definite matrices of rank higher than r but not against general matrices of rank higher than r. That is, the t statistic in Example 2 diverges under $H_1(r)$ not just because $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r} = O_p^{-1}(1)$ but also because $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r}$ approaches $\mathcal{P}=\mathbb{P}^m$.

All of the statistics we have considered so far satisfy Assumptions K (see Table 2). Robin & Smith (2000) take $\kappa(X,Y)=\lVert Y^{\dagger/2}\text{vec}(X)\rVert^2+O(\lVert Y^{\dagger/2}\text{vec}(X)\rVert^3)$ as $Y^{\dagger/2}\text{vec}(X)\to 0$. This class of statistics is asymptotically equivalent to the class of statistics satisfying Assumptions K with $\theta=1$, $\mathcal{P}=\mathcal{X}=\mathbb{R}^{n\times m}$, and $\mathcal{Y}=\mathbb{P}^{nm}$. Another option is to take any norm φ on $\mathbb{R}^{nm}$ and formulate a statistic with $\kappa(X,Y)=\varphi^2(Y^{\dagger/2}\text{vec}(X))$ (the F statistic is a special case of this). Here, $\theta=1$, $\mathcal{P}=\mathcal{X}=\mathbb{R}^{n\times m}$, and $\mathcal{Y}=\mathbb{P}^{nm}$.

We can now state the first main theorem of the paper.

Theorem 1. Suppose Assumptions K hold along with either Assumptions A or B. Suppose the null space estimators $\widehat{N}_r\in\mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r\in\mathbb{G}^{m\times(m-r)}$ are obtained by either a DBA or a CDA with $\text{cond}(\Theta)=O_p(1)$. Let $N_r\in\mathbb{G}^{n\times(n-r)}$ and $M_r\in\mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $(B^*)_r^{RRA}$ for $r\le\text{rank}(B^*)$. Suppose (i) the following inclusions hold almost surely

$$P_{N_r}\widehat{B}P_{M_r}\in\mathcal{X}, \qquad (P_{M_r}\otimes P_{N_r})\widehat{\Omega}(P_{M_r}\otimes P_{N_r})\in\mathcal{Y},$$

$$P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}\in\mathcal{X}, \qquad (P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\in\mathcal{Y},$$

and (ii) the following conditions hold

$$d(P_{N_r}\widehat{B}P_{M_r}, \mathcal{P}) = o_p(1), \qquad\qquad d(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}, \mathcal{P}) = o_p(1).$$

Then $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) = T^\theta \kappa \left(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}, (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) \right)$ satisfies the weak plug–in principle for statistics of tests of rank. If, additionally, the RRA is continuous at B^*, then the statistic satisfies the strong plug–in principle.

Theorem 1 requires two additional assumptions: (i) that the feasible and infeasible statistics be well defined and (ii) that the terms $P_{N_r}\widehat{B}P_{M_r}$ and $P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}$ approach the region where the statistics can have power. Condition (ii) is automatically satisfied for statistics in which $\mathcal{X} = \mathcal{P}$ (see Table 2). The intuition behind Theorem 1 is as follows. Under $H_0(r)$ or $H_T(r)$, we know from Lemma 1 (i) that $\widehat{N}_r$ and $\widehat{M}_r$ are consistent null space estimators and this together with condition (i) of Assumptions K (i.e. smoothness), implies that substituting the null space estimators for their limiting values affects negligible change to the value of the statistic; this satisfies condition (i) of the plug–in principle. Under $H_1(r)$, we know by Lemma 1 (ii) that $P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}$ is bounded away from zero in probability and if it additionally approaches $\mathcal{P}$ then condition (ii) of Assumptions K implies that $\kappa \left(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}, (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) \right)$ will be bounded away from zero in probability; this satisfies condition (ii) of the plug–in principle. Finally, under $H_1(r)$, when the RRA is continuous at B^*, then we know from Lemma 1 (iii) that the null space estimators will converge to the null spaces of $(B^*)_r^{RRA}$ and the smoothness of κ again implies that the substitution of the null space estimators for their population analogues has negligible effect on the rate of divergence of the statistic under $H_1(r)$; this satisfies condition (iii) of the plug–in principle.

It follows from Theorem 1 that all of the statistics of the standard asymptotics literature (and the many more we have listed above) satisfy the weak plug–in principle; when the underlying RRA is continuous at the population matrix (a feature we have conjectured to be generic for all RRAs), they also satisfy the strong plug–in principle.

It also follows from Theorem 1 that statistics with common functional form but different null space estimators differ from each other by $O_p(T^{-1/2})$ under $H_0(r)$ and $H_T(r)$. Thus, the Cragg & Donald (1996), Cragg & Donald (1997), and Kleibergen & Paap (2006) statistics do not only have the same limiting distribution under $H_0(r)$ and $H_T(r)$, but they differ from each other by $O_p(T^{-1/2})$. When $\widehat{\Omega}$ is of Kronecker product form, we may add to the list the statistics of Anderson (1951), Robin & Smith (1995), and Robin & Smith (2000). In the symmetric case, we have additionally that all three F statistics proposed by Donald et al.

(2007) differ from each other by $O_p(T^{-1/2})$ under $H_0(r)$ or $H_T(r)$. In choosing between these equivalent statistics then, the practitioner must rely on either small sample performance or numerical convenience in choosing the right test. In the latter case, we note that the CDA with non–Kronecker product weighting matrix is the most computationally expensive of the RRAs considered in this paper. Next are the RSD, SVD, and EIG RRAs, which although much faster than the CDA, are not the most efficient computationally. The fastest available algorithms are the LU and QR algorithms (Hansen, 1998; Golub & Van Loan, 1996). Therefore, these latter algorithms are recommended for high intensity computations such as the bootstrap.

An immediate corollary of Theorem 1 is that the test for identification proposed by Wright (2003) does not have to be conducted using the Cragg & Donald (1997) statistic but can instead be done using any other asymptotically equivalent F statistic. The same statistic can also be avoided in the rank estimator proposed by Cragg & Donald (1997).

The weak plug–in principle simplifies the asymptotics of tests of rank tremendously. It allows us to immediately see the asymptotic distribution under $H_0(r)$ and $H_T(r)$ – we simply derive the asymptotic distribution as if the population null spaces were known. It also allows us to obtain the asymptotics under $H_1(r)$ and misspecification. The strong plug–in principle, in turn, allows us (under possibly generic conditions) to obtain precise estimates of the rates of divergence of the statistics under $H_1(r)$. See Section 5 for Monte Carlo illustrations of the weak and strong plug–in principles in standard asymptotics.

Note that we may relax the condition that $\widehat{\Omega}^{-1} = O_p(1)$ (resp. $\widehat{\Psi}^{-1} = O_p(1)$) under Assumptions A (resp. B). Here, there are two cases to consider: reducible singularity, which can be treated by rescaling (this is taken up in the next section), and irreducible singularity, which requires regularization (the terminology is due to Dufour & Valery (2011)). In the latter case, we may pursue the approach of Moore (1977) if we can ensure that $(P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})$ satisfies the conditions of Andrews (1987). If not, we will need to substitute $Y^{\dagger}$ in Table 2 with one of the regularized inverses proposed by Lütkepohl & Burda (1997) or Dufour & Valery (2011). Recently, Duplinskiy (2014) has proposed avoiding regularization altogether and simply bootstrapping the non–standardized test statistics. Donald et al. (2014) consider this problem in greater detail.

As an application of the plug–in principle, we explicitly derive the limiting distributions of some of the statistics we have considered above.

Corollary 1. Suppose Assumptions K and A hold and suppose we have null space estimators $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ obtained by either a DBA or a CDA with

$\text{cond}(\Theta) = O_p(1)$. Under $H_0(r)$ or $H_T(r)$, let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of B^* respectively. Then if $T^\theta\tau\left(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r}\right) \xrightarrow{d} \zeta$, then $T^\theta\tau\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \zeta$. In particular, if

$$\left(\sqrt{T}\text{vec}(N_r'\widehat{B}M_r), (M_r \otimes N_r)'\widehat{\Omega}(M_r \otimes N_r)\right) \xrightarrow{d} (\xi_r, \Omega_r),$$

then we have

$$LR_g\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\xi_r\|^2_{\Omega_r}, \qquad F\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\xi_r\|^2_{\Omega_r},$$
$$J_g\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\text{mat}(\Omega_r^{-1/2}\xi_r)\|_2^2.$$

Corollary 1 implies that, if $\sqrt{T}\text{vec}(\widehat{B} - B) \xrightarrow{d} N(0, \Phi)$ and $\widehat{\Omega}$ converges to a constant positive definite matrix, then LR_g and F converge in distribution to a quadratic form in $(n-r)(m-r)$ normal random variables, while J_g converges to the square of the 2–norm of a random matrix with normal entries. Under correct specification (i.e. $\widehat{\Omega} \xrightarrow{p} \Phi$) and $H_0(r)$, $F \xrightarrow{d} \chi^2((n-r)(m-r))$ and $J_g \xrightarrow{d} \|Z\|_2^2$, where $\text{vec}(Z) \sim N(0, I_{(n-r)(m-r)})$. Under correct specification and $H_T(r)$, $F \xrightarrow{d} \chi^2\left((n-r)(m-r), \|N_r'DM_r\|^2_{(M_r\otimes N_r)'\Phi(M_r\otimes N_r)}\right)$ and $J_g \xrightarrow{d} \left\|Z + \text{mat}\left(((M_r \otimes N_r)'\Phi(M_r \otimes N_r))^{\dagger/2}\text{vec}(N_r'DM_r)\right)\right\|_2^2$, with Z as before. For the limiting distribution of LR_g and F under $H_0(r)$ and $H_T(r)$ and incorrect specification (i.e. $\widehat{\Omega} \xrightarrow{p} \Omega \neq \Phi$), the reader is referred to Lemma 8.2 of White (1994). The limiting distributions of J_g under $H_0(r)$ and $H_T(r)$ and incorrect specification is non–standard and does not appear to simplify further than what is stated in the result above. Under fixed–b asymptotics $\widehat{\Omega}$ does not converge in probability although, under the usual assumptions, the limiting distributions of the statistics above are free of nuisance parameters. The reader is referred to the fixed–b literature for the limiting distributions (Kiefer et al., 2000; Kiefer & Vogelsang, 2002a,b, 2005).

Corollary 1 generalizes the misspecification results of Robin & Smith (2000), who consider the asymptotics of only the A statistic. It also generalizes the local power result of Cragg & Donald (1997), who consider only the F statistic that employs the CDA null space estimator. Finally, it allows for more general functional forms of κ than previously used in the literature.

Corollary 2. Suppose Assumptions K and B hold and suppose we have a null space estimator $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ obtained by either a DBA or a CDA with $\text{cond}(\Theta) = O_p(1)$. Under $H_0(r)$ or $H_T(r)$, let $M_r \in \mathbb{G}^{m\times(m-r)}$ span the null space of B^*. Then if $T^\theta\tau\left(\widehat{B}, \widehat{\Omega}, P_{M_r}, P_{M_r}\right) \xrightarrow{d} \zeta$, then $T^\theta\tau\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \zeta$. In particular, if

$$\left(\sqrt{T}\text{vech}(M_r'\widehat{B}M_r), D_{m-r}^\dagger(M_r \otimes M_r)'\widehat{\Omega}(M_r \otimes M_r)D_{m-r}^{\dagger\prime}\right) \xrightarrow{d} (\xi_r, \Omega_r),$$

then we have

$$LR_g\left(\widehat{B},\widehat{\Omega},P_{\widehat{M}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\xi_r\|^2_{\Omega_r}, \qquad F\left(\widehat{B},\widehat{\Omega},P_{\widehat{M}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\xi_r\|^2_{\Omega_r},$$
$$J_g\left(\widehat{B},\widehat{\Omega},P_{\widehat{M}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\text{mat}(D_{m-r}\Omega_r^{-1/2}\xi_r)\|_2^2,$$

and if M_r is chosen to have orthogonal columns then

$$t\left(\widehat{B},\widehat{\Psi},P_{\widehat{M}_r}\right)\xrightarrow{d}\frac{\text{tr}(\text{mat}(D_{m-r}\xi_r))}{(\text{vec}'(I_{m-r})D_{m-r}\Omega_r D'_{m-r}\text{vec}(I_{m-r}))^{1/2}}.$$

It follows from Corollary 2 that if $\sqrt{T}\text{vech}(\widehat{B}-B)\xrightarrow{d}N(0,\Phi)$ and $\widehat{\Psi}$ converges in probability to a positive definite matrix, then LR_g and F converge in distribution to a quadratic form in $(m-r)(m-r+1)/2$ normal random variables, J_g converges to the square of the 2–norm of a random matrix with normal entries, and t converges to a normal random variable. Under correct specification (i.e. $\widehat{\Psi}\xrightarrow{p}\Phi$) and $H_0(r)$, $F\xrightarrow{d}\chi^2((m-r)(m-r+1)/2)$, $J_g\xrightarrow{d}\|Z\|_2^2$, where $Z=Z'$ and $\text{vech}(Z)\sim N(0,I_{(m-r)(m-r+1)/2})$, and $t\xrightarrow{d}N(0,1)$. Under correct specification and $H_T(r)$, $F\xrightarrow{d}\chi^2\left((m-r)(m-r+1)/2,\|\text{vech}(M_r'DM_r)\|^2_{D^{\dagger}_{m-r}(M_r\otimes M_r)'\Phi(M_r\otimes M_r)D^{\dagger}_{m-r}{}'}\right)$, $J_g\xrightarrow{d}\left\|Z+\text{mat}(D_{m-r}((D^{\dagger}_{m-r}(M_r\otimes M_r)'\Phi(M_r\otimes M_r)D^{\dagger}_{m-r}{}')^{-1/2}\text{vech}(M_r'DM_r)))\right\|_2^2$, with Z the same as before, and $t\xrightarrow{d}N\left(\frac{\text{tr}(M_r'DM_r)}{\text{vec}'(I_{m-r})(M_r\otimes M_r)'\Phi(M_r\otimes M_r)\text{vec}(I_{m-r})},1\right)$. For the limiting distribution of F and t under $H_0(r)$ and $H_T(r)$ and incorrect specification (i.e. $\widehat{\Psi}\xrightarrow{p}\Psi\neq\Phi$), the reader is referred again to Lemma 8.2 of White (1994). The limiting distributions of J_g under $H_0(r)$ and $H_T(r)$ and incorrect specification is, again, not amenable to further simplification. Fixed–b asymptotics are treated in the literature cited following Corollary 1.

Donald et al. (2007) proved the $H_0(r)$ and $H_1(r)$ results for F and t in the case of correct specification. Thus Corollary 2 extends their results in the direction of local power, misspecification, fixed–b asymptotics, and more general functional forms of the statistics.

4 Tests of Rank Under Non–Standard Asymptotics

In this section, we will extend the basic theory to more general settings that allow for cointegration. Just as before, we consider some well–known examples in the literature before proceeding to generalize null space estimation and the plug–in principle.

4.1 Preliminary Examples

Cointegration presents some truly fascinating anomalies for tests of rank. In the next couple of examples we will show that the framework of Section 3 cannot be applied verbatim. The examples will, however, point to the necessary generalization.

Example 3. Let $\{\varepsilon_t : t \geq 1\}$ be i.i.d. $N(0, \Sigma)$, $\Sigma \in \mathbb{P}_+^m$, $y_0 = 0$, and

$$\Delta y_t = By_{t-1} + \varepsilon_t, \qquad t = 1, \dots, T.$$

We assume that the roots of the characteristic polynomial of the system are either outside the unit circle or else at 1. Assume for the moment that the model generates data of order of integration no higher than 1 (see Theorem 4.2 of Johansen (1995a) for the conditions). Then $r = \text{rank}(B) < m$ is the number of cointegration relationships. Let $N_r \in \mathbb{G}^{m \times (m-r)}$ and $M_r \in \mathbb{G}^{m \times (m-r)}$ span the left and right null spaces of B respectively. Let $\widehat{B}$ and $\widehat{\Sigma}$ be the OLS estimators of B and Σ and let $\widehat{\Omega} = \widehat{\Gamma}^{-1} \otimes \widehat{\Sigma}$, where $\widehat{\Gamma} = \frac{1}{T} \sum_{t=1}^{T} y_{t-1} y_{t-1}'$.

It is easy to check that Johansen's (1988) trace statistic is $LR\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right)$, where the null space estimators are the RSD estimators, and has the same limiting distribution as that of $LR\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{N_r}, P_{M_r}\right)$. This suggests that the plug–in principle holds here. Unfortunately, however, $\widehat{\Omega}$ converges to a singular matrix so Assumptions A fail. On closer inspection, however, we find that $\widehat{\Omega}$'s rate of convergence along its asymptotic null space is counterbalanced by $\widehat{B}$'s rate of convergence along its asymptotic right null space. In particular, $(T^{1/2} M_r \otimes I_n)' \widehat{\Omega} (T^{1/2} M_r \otimes I_n)$ and $T\widehat{B} M_r$ are each $O_p(1)$. This counterbalancing of the accelerated rates of convergence is of crucial importance in the theory of cointegration.

Now suppose that the order of integration is no higher than 2 (see Theorem 4.6 of Johansen (1995a) for the conditions). Then Johansen (1995b) finds that $M_r = [\ M_{r1} \quad M_{r2}\ ]$, where $\left([\ T^{1/2} M_{r1} \quad T^{3/2} M_{r2}\ ]' \otimes I_n\right) \widehat{\Omega} \left([\ T^{1/2} M_{r1} \quad T^{3/2} M_{r2}\ ] \otimes I_n\right)$ and $\widehat{B}[\ T M_{r1} \quad T^2 M_{r2}\ ]$ are $O_p(1)$. Thus, there may be heterogenous rates of accelerated convergence that need to be taken into account. □

The phenomenon illustrated in Example 3 is well known in cointegration (Johansen, 1995a) and in regressions with polynomial trends (Hamilton, 1994, Chapter 16). $\widehat{B}$ and $\widehat{\Omega}$ shrink to zero along certain directions at offsetting rates. Thus, all that is required to evaluate the asymptotics of the infeasible statistic is to rescale $\widehat{B}$ and $\widehat{\Omega}$ by the appropriate power of T along the appropriate directions (provided the statistic is invariant to such rescaling). For this reason, Dufour & Valery (2011) refer to the limiting singularity of $\widehat{\Omega}$ as reducible singularity. We will show that if the statistic is invariant to such rescaling, then the feasible statistic continues to mimic the infeasible statistic, thus proving the plug–in principle.[5] However, this is not the only subtlety in cointegration rank testing as the next example shows.

[5]In fact, it will be seen that the rates found in Example 3 are stronger than necessary for satisfying the coun-

Example 4. Let $\{(\varepsilon_t', u_t')' : t \geq 1\}$ be a $2m$–dimensional sequence of i.i.d. $N\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}\Sigma & 0\\0 & B\end{smallmatrix}\right]\right)$ random vectors, $\Sigma \in \mathbb{P}_+^m$, $x_0 = 0$, and

$$y_t = x_t + \varepsilon_t, \qquad x_t = x_{t-1} + u_t, \qquad t = 1, \ldots, T.$$

The rank of $B \in \mathbb{R}^{m\times m}$ determines the number of stochastic trends in the model. Let $M_r \in \mathbb{G}^{m\times(m-r)}$ span the null space of B. Let $\overline{y} = T^{-1}\sum_{t=1}^T y_t$, $\widehat{\Sigma} = T^{-2}\sum_{t=1}^T (y_t - \overline{y})(y_t - \overline{y})' \in \mathbb{P}_+^m$, and $\widehat{\Gamma} = T^{-4}\sum_{t=1}^T \sum_{s=1}^t (y_s - \overline{y}) \sum_{s=1}^t (y_s - \overline{y})'$. We will work with $\widehat{B} = \widehat{\Sigma}^{-1/2}\widehat{\Gamma}\widehat{\Sigma}^{-1/2}$.

Nyblom & Harvey (2000) show that $\widehat{B}$ converges in distribution to a random matrix whose null space is exactly the span of M_r. As $[\ M_{r\perp} \quad \sqrt{T}M_r\ ]'\,\widehat{B}\,[\ M_{r\perp} \quad \sqrt{T}M_r\ ]$ converges in distribution to an almost surely positive definite matrix, they propose the t–type statistic $T\mathrm{tr}(P_{\widehat{M_r}}\widehat{B}P_{\widehat{M_r}})$, where $\widehat{M_r}$ is the EIG null space estimator based on $\widehat{B}$. Experience would then suggest that this statistic should mimic $T\mathrm{tr}(P_{M_r}\widehat{B}P_{M_r})$. Surprisingly, however, this is not the case. It would seem then that the plug–in principle fails.

In fact, the plug–in principle still holds but for a different matrix than M_r. One can check that $\widehat{B}M_r = O_p(T^{-1/2})$. On the other hand, the Poincaré Separation Theorem implies that the smallest $m - r$ eigenvalues of $\widehat{B}$ are $O_p(T^{-1})$. Thus, normalizing and collecting the associated eigenvectors in $M_{rT} \in \mathbb{G}^{m\times(m-r)}$, we have that $\widehat{B}M_{rT} = O_p(T^{-1})$. Therefore, we find the surprising fact that M_r fails to capture the appropriate rate of convergence of $\widehat{B}$ to singularity and there are other directions along which $\widehat{B}$ converges faster. Another example, based on a deeper analysis of the asymptotics of $\widehat{B}$, is $M_{rT} = \sqrt{T}\widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M_{r\perp}'\widehat{\Gamma}M_{r\perp})^{-1}M_{r\perp}'\widehat{\Gamma})M_r(M_r'\Sigma M_r)^{-1/2} = M_r + o_p(1)$ and satisfies $\widehat{B}M_{rT} = O_p(T^{-1})$. In both cases, $\widehat{B}$ converges faster along M_{rT} than it does along M_r, even though $P_{M_{rT}}$ converges to P_{M_r}. The crucial point to note here is that any reasonable null space estimators will detect M_{rT} rather than M_r. Thus, the plug–in principle continues to hold, albeit for M_{rT} rather than M_r and the limiting distribution of $T\mathrm{tr}(P_{M_{rT}}\widehat{B}P_{M_{rT}})$ is precisely the limiting distribution of the Nyblom and Harvey statistic. See on-line appendix E.4 for the mathematical details. $\square$

Example 4 suggests three additional subtleties. First, $\widehat{B}$ need not converge in probability even as some linear combinations of its rows and columns converge to zero in probability. Second, rescaling should be allowed along possibly random and T–varying directions. Third, statistics in cointegration may be scaled differently than in standard asymptotics as T appears with a power of 1 instead of $\frac{1}{2}$ in the Nyblom & Harvey (2000) statistic. The appropriate scaling factor will be explained below.

terbalancing assumption we will need for the plug–in principle. See on-line appendix E.3 for further details on the $I(1)$ case.

Using these statistics, we will be interested in testing $H_0(r)$ against $H_1(r)$ and $H_T(r)$, which is now defined as

$$H_T(r) : B = B^* + T^{-\omega} D, \quad \text{rank}(B^*) = r,$$

where $\omega \geq 0$ ensuring that the local alternative does not stray too far away from B^*. In Example 3 the appropriate ω is 1 for the $I(1)$ case and 2 for the $I(2)$ case (Hubrich et al., 2001). In Example 4, the appropriate ω is 2. See on-line appendices E.3 and E.4.[6]

4.2 Estimating the Null Spaces

In this section we consider the performance of null space estimators when $\widehat{B}$ converges to zero along non–constant directions at heterogenous rates. The main idea remains the same, $\widehat{B}$ will be seen to approach a sequence of reduced–rank matrices $\widehat{B}^*$ whose limiting null spaces we would like to estimate. If $\widehat{B}_r^{RRA}$ approximates $\widehat{B}$ well enough, then it will approximate $\widehat{B}^*$ as well and we may estimate the null spaces of interest by the null spaces of $\widehat{B}_r^{RRA}$. Thus, $\widehat{B}^*$ plays the role that B^* played in Section 3.2.

Lemma 2. Let $\widehat{B}$ be an estimator indexed by T such that $\widehat{B} \in \mathbb{G}^{n\times m}$ almost surely and $\widehat{B} = O_p(1)$. Let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ and suppose there exists sequences of possibly random matrices $N_{rT} \in \mathbb{G}^{n\times(n-r)}$ and $M_{rT} \in \mathbb{G}^{m\times(m-r)}$, whose singular values are bounded away from zero in probability, $P_{N_{rT}} \xrightarrow{p} P_{N_r}$ and $P_{M_{rT}} \xrightarrow{p} P_{M_r}$, and, for $\gamma > 0$,

$$\sigma_r(N_{r\perp}' \widehat{B} M_{r\perp}) = O_p^{-1}(1), \quad T^\gamma N_{rT}' \widehat{B} = O_p(1), \quad T^\gamma \widehat{B} M_{rT} = O_p(1), \quad T^\gamma N_{rT}' \widehat{B} M_{rT} = O_p(1).$$

Let the RRAs $\{\widehat{B}_i^{RRA} : 0 \leq i < \min\{n, m\}\}$ be either DBAs or CDAs. In the latter case, we assume that $\text{cond}(\Theta_T) = O_p(1)$, where $\Theta_T = Z_T' \Theta Z_T$ and $Z_T = [\, M_{r\perp} \quad M_{rT} \,] \otimes [\, N_{r\perp} \quad N_{rT} \,]$. Let $\widehat{N}_i$ (resp. $\widehat{M}_i$) span the left (resp. right) null space of $\widehat{B}_i^{RRA}$ for $0 \leq i < \min\{n, m\}$. Finally, set $\widehat{B}^* = [\, N_{r\perp} \quad N_{rT} \,]^{-1\prime} \begin{bmatrix} N_{r\perp}' \widehat{B} M_{r\perp} & 0 \\ 0 & 0 \end{bmatrix} [\, M_{r\perp} \quad M_{rT} \,]^{-1}$.

(i) $T^\gamma(\widehat{B} - \widehat{B}^*)$, $T^\gamma(\widehat{B} - \widehat{B}_r^{RRA})$, $T^\gamma(P_{\widehat{N}_r} - P_{N_{rT}})[\, N_{r\perp} \quad N_{rT} \,]$, and $T^\gamma(P_{\widehat{M}_r} - P_{M_{rT}})[\, M_{r\perp} \quad M_{rT} \,]$ are $O_p(1)$.

(ii) If $0 \leq i < r$ and the null spaces are estimated by DBA, then $P_{\widehat{N}_i} \widehat{B} P_{\widehat{M}_i} = O_p^{-1}(1)$. If, on the other hand, the null spaces are estimated by CDA, then $P_{\widehat{N}_{iT}} \widehat{B}_T P_{\widehat{M}_{iT}} = O_p^{-1}(1)$ and $[\, N_{r\perp} \quad N_{rT} \,]' P_{\widehat{N}_i} \widehat{B} P_{\widehat{M}_i} [\, M_{r\perp} \quad M_{rT} \,] = O_p^{-1}(1)$, where $\widehat{N}_{iT} = [\, N_{r\perp} \quad N_{rT} \,]^{-1} \widehat{N}_i$, $\widehat{M}_{iT} = [\, M_{r\perp} \quad M_{rT} \,]^{-1} \widehat{M}_i$, and $\widehat{B}_T = [\, N_{r\perp} \quad N_{rT} \,]' \widehat{B} [\, M_{r\perp} \quad M_{rT} \,]$.

[6] Of course, more complicated local alternatives can also be considered (e.g. Hallin et al. (2016)). Since this is highly contingent on the DGP, we opt to simplify the exposition by considering only the simplest case.

If $n = m$, $N_{r\perp} = M_{r\perp}$, $N_{rT} = M_{rT}$, and $d(M'_{r\perp}\widehat{B}M_{r\perp}, \mathbb{P}^r) = o_p(1)$, then $P_{\widehat{M}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ and $P_{\widehat{N}_i}\widehat{B}P_{\widehat{N}_i} = O_p^{-1}(1)$ for the DBA and $P_{\widehat{M}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}} = O_p^{-1}(1)$ and $P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{N}_{iT}} = O_p^{-1}(1)$ for the CDA.

(iii) If $0 \leq i < r$ and the distance between $\widehat{B}^*$ and the set of discontinuity points of the rank–i RRA is bounded away from zero in probability, then $P_{\widehat{N}_i} - P_{\widehat{N}_i^*} = o_p(1)$ and $P_{\widehat{M}_i} - P_{\widehat{M}_i^*} = o_p(1)$, where $\widehat{N}_i^*$ and $\widehat{M}_i^*$ span the left and right null spaces of $(\widehat{B}^*)_i^{RRA}$ respectively.

The assumptions of Lemma 2 specialize to those of Lemma 1 when $N_{rT} = N_r$ and $M_{rT} = M_r$ span the null spaces of some matrix B^* and $\sqrt{T}(\widehat{B} - B^*) = O_p(1)$ so that $\gamma = \frac{1}{2}$ and $\widehat{B}^* = B^*$. These assumptions accommodate the features we have documented in Examples 3 and 4. First, $\widehat{B}$ is allowed to converge along N_{rT} and M_{rT} at any non–standard rate γ and, since they both depend on T, heterogeneous rates are also allowed. In the $I(1)$ case of Example 3, $\gamma = \frac{1}{2}$, $N_{rT} = N_r$, and $M_{rT} = T^{1/2}M_r$, while in the $I(2)$ case $M_{rT} = [\ T^{1/2}M_{r1} \quad T^{3/2}M_{r2}\ ]$. In this regard, the condition that N_{rT} and M_{rT} have singular values bounded away from zero in probability is important in order to to ensure that they specify proper directions along which $\widehat{B}$ goes to zero.[7] Second, the matrices N_{rT} and M_{rT} may be random, allowing for the setting we had in Example 4, where $\gamma = 1$ and $N_{rT} = M_{rT}$ were chosen as $\sqrt{T}\widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma})M_r(M'_r\Sigma M_r)^{-1/2}$. Third, $\widehat{B}$ is only required to converge in probability along N_{rT} and M_{rT} but its components along $N_{r\perp}$ and $M_{r\perp}$ are only required to be bounded in probability. This allows for the scenario we witnessed in Example 4, where $\widehat{B}$ converges in distribution, but not in probability, along $N_{r\perp}$ and $M_{r\perp}$. In this regard, the boundedness of $\sigma_r(N'_{r\perp}\widehat{B}M_{r\perp})$ away from zero in probability is necessary in order to ensure $\widehat{B}$ has an $r \times r$ asymptotically surviving component even as its components along N_{rT} and M_{rT} are vanishing. When $\widehat{B}$ converges to a fixed rank–r matrix, as in Example 3 or Section 3, this boundedness condition is redundant.

It is important to emphasize that γ, the rate of convergence of $\widehat{B}$ along N_{rT} and M_{rT}, need not be restricted to integer multiples of $\frac{1}{2}$ but can take any positive value in general. For cointegrated stable processes (Caner, 1998) and fractionally cointegrated processes (Johansen & Nielsen, 2012), for example, γ can take values in the interval $\left(\frac{1}{2}, \infty\right)$.

Lemma 2 (i) proves that even though $\widehat{B}$ is not ensured to converge, it does approach a random sequence $\widehat{B}^*$ of matrices of rank–r at a rate of $O_p(T^{-\gamma})$. The rates of convergence of

[7] Otherwise, $\widehat{B}M_{rT} \xrightarrow{p} 0$ could occur not because $\widehat{B}$ converges to zero in probability along M_{rT} but because certain line combinations of the columns of M_{rT} converge to zero in probability.

DBAs and CDAs to $\widehat{B}^*$ are also found to be $O_p(T^{-\gamma})$. The rates of convergence of the null space estimators in Lemma 2 (i) may seem peculiar compared to their counterparts in Lemma 1 (i). In fact, the expressions are more parsed descriptions of the convergence of null space estimators in two important respects. First, they allow for the estimated null spaces to exhibit accelerated rates of convergence. For example, the lemma implies that the right null space estimator in the $I(1)$ case of Example 3 satisfies $T(P_{\widehat{M}_r}M_r - M_r) = O_p(1)$ and, utilizing a similar argument to that used in the proof of the lemma, this then implies that $T(P_{\widehat{M}_r} - P_{M_r}) = O_p(1)$. The left null space estimator, on the other hand, has a standard rate of convergence $\sqrt{T}(P_{\widehat{N}_r} - P_{N_r}) = O_p(1)$. Second, they allow subspaces of the estimated null spaces to exhibit heterogeneous rates of convergence. In the $I(2)$ case of Example 3, $T(P_{\widehat{M}_r}M_{r1} - M_{r1}) = O_p(1)$ and $T^2(P_{\widehat{M}_r}M_{r2} - M_{r2}) = O_p(1)$. Thus, it is possible to decompose $\widehat{M_r}$ as $[\ \widehat{M_{r1}}\quad \widehat{M_{r2}}\ ]$ with $T(P_{\widehat{M}_{r1}} - P_{M_{r1}}) = O_p(1)$ and $T^2(P_{\widehat{M}_{r2}} - P_{M_{r2}}) = O_p(1)$. Geometrically, the canonical angles between $\text{span}(\widehat{M_r})$ and $\text{span}(M_r)$ converge to zero at different rates. Note that when the assumptions of Lemma 2 specialize to those of Lemma 1, the new expressions for the rates of convergence of the null space estimators are equivalent to the old ones.

Lemma 2 (ii) provides the same results as Lemma 1 (ii) for the DBA but not for the CDA. For the CDA, we are only ensured the weaker result that $P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}} = O_p^{-1}(1)$, which implies the even weaker result that $[\ N_{r\perp}\quad N_{rT}\ ]' P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}[\ M_{r\perp}\quad M_{rT}\ ] = O_p^{-1}(1)$. That is, we are only ensured that a rescaled version of $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}$ is bounded away from zero in probability. However, when N_{rT} and M_{rT} are bounded in probability we still have that $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}$ is bounded away from zero in probability and when N_{rT} and M_{rT} are unbounded, we will see that the CDA is still capable of delivering power in tests of rank based on statistics invariant to rescaling.

Lemma 2 (iii) states that when $\widehat{B}^*$ remains well within the region of continuity of the RRA, then as $\widehat{B}$ approaches this sequence the associated null space estimators converge in probability to the null spaces of $(\widehat{B}^*)_i^{RRA}$. It follows that $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_r} - P_{\widehat{N}_i^*}\widehat{B}P_{\widehat{M}_r^*} = o_p(1)$ and we will again have an estimator of the non–vanishing components of $\widehat{B}$.

4.3 The Plug–In Principle

Consider the following generalizations of Assumptions A and B.

Assumptions C. $B^* \in \mathbb{R}^{n\times m}$. $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{S}^{nm}$ are estimators indexed by T. Each $\text{vec}(\widehat{B}) \in \mathbb{R}^{nm}$ is a non–degenerate random vector and $\widehat{B} = O_p(1)$. $\widehat{\Omega} \in \mathbb{P}_+^{nm}$ almost surely. If $N_q \in \mathbb{G}^{n\times(n-q)}$ and $M_q \in \mathbb{G}^{m\times(m-q)}$ span the left and right null spaces of B^* respectively, there exists sequences of possibly random matrices $N_{qT} \in \mathbb{G}^{n\times(n-q)}$ and $M_{qT} \in \mathbb{G}^{m\times(m-q)}$, whose singular values are bounded away from zero in probability, $P_{N_{qT}} \xrightarrow{p} P_{N_q}$ and $P_{M_{qT}} \xrightarrow{p} P_{M_q}$, and, for $\gamma > 0$,

$$\sigma_q(N'_{q\perp}\widehat{B}M_{q\perp}) = O_p^{-1}(1), \qquad T^\gamma N'_{qT}\widehat{B} = O_p(1), \qquad T^\gamma \widehat{B}M_{qT} = O_p(1),$$
$$T^\gamma N'_{qT}\widehat{B}M_{qT} = O_p(1), \qquad Z'_T\widehat{\Omega}Z_T = O_p(1), \qquad (Z'_T\widehat{\Omega}Z_T)^{-1} = O_p(1),$$

where $Z_T = [\ M_{q\perp} \quad M_{qT}\ ] \otimes [\ N_{q\perp} \quad N_{qT}\ ]$.

The symmetric analogue is given by the following set of assumptions.

Assumptions D. $B^* \in \mathbb{S}^m$. $\widehat{B} \in \mathbb{S}^m$ and $\widehat{\Psi} \in \mathbb{S}^{m(m+1)/2}$ are estimators indexed by T. Each $\mathrm{vec}(\widehat{B}) \in \mathbb{R}^{m(m+1)/2}$ is a non–degenerate random vector and $\widehat{B} = O_p(1)$. $\widehat{\Psi} \in \mathbb{P}_+^{m(m+1)/2}$ almost surely. If $M_q \in \mathbb{G}^{m\times(m-q)}$ spans the null space of B^*, there exists a sequence of possibly random matrices $M_{qT} \in \mathbb{G}^{m\times(m-q)}$, whose singular values are bounded away from zero in probability, $P_{M_{qT}} \xrightarrow{p} P_{M_q}$, and, for $\gamma > 0$,

$$\sigma_q(M'_{q\perp}\widehat{B}M_{q\perp}) = O_p^{-1}(1), \qquad T^\gamma \widehat{B}M_{qT} = O_p(1), \qquad T^\gamma M'_{qT}\widehat{B}M_{qT} = O_p(1),$$
$$D_m^\dagger Z'_T D_m \widehat{\Psi} D'_m Z_T D_m^{\dagger\prime} = O_p(1), \qquad (D_m^\dagger Z'_T D_m \widehat{\Psi} D'_m Z_T D_m^{\dagger\prime})^{-1} = O_p(1),$$

where $Z_T = [\ M_{q\perp} \quad M_{qT}\ ] \otimes [\ M_{q\perp} \quad M_{qT}\ ]$. In this context, we will set $\widehat{\Omega} = D_m\widehat{\Psi}D'_m$.

The alternatives $H_0(r)$, $H_T(r)$, and $H_1(r)$ interact with Assumptions C and D as they did before, by setting $q = \mathrm{rank}(B^*)$. In particular, $q = r$ under $H_0(r)$ or $H_T(r)$ and $q > r$ under $H_1(r)$. However, unlike in Assumptions A and B, q appears explicitly because the asymptotics in cointegration are intimately related to the rank of B^*.

Clearly, Assumptions C and D reduce to Assumptions A and B respectively, when $\sqrt{T}(\widehat{B} - B^*) = O_p(1)$ and N_{qT} and M_{qT} are independent of T. These assumptions establish sufficient conditions for Lemma 2 to hold under the non–symmetric and symmetric settings respectively. The only additional assumptions concern the asymptotic behaviour of $\widehat{\Omega}$ and $\widehat{\Psi}$. These are required to be asymptotically bounded and well–conditioned after conformable rescaling with $\widehat{B}$. This is the manifestation of the counterbalancing effect we discussed in Example 3. In the $I(1)$ case of Example 3, we need $\Big([\ M_{q\perp} \quad T^{1/2}M_q\ ]'\widehat{\Gamma}^{-1}[\ M_{q\perp} \quad T^{1/2}M_q\ ]\Big) \otimes \Big([\ N_{q\perp} \quad N_q\ ]'\widehat{\Sigma}[\ N_{q\perp} \quad N_q\ ]\Big)$ and its inverse to be $O_p(1)$, while in the $I(2)$ case we need $\Big([\ M_{q\perp}\ \ T^{1/2}M_{q1}\ \ T^{3/2}M_{q2}\ ]'\widehat{\Gamma}^{-1}[\ M_{q\perp}\ \ T^{1/2}M_{q1}\ \ T^{3/2}M_{q2}\ ]\Big) \otimes \Big([\ N_{q\perp}\ \ N_q\ ]'\widehat{\Sigma}[\ N_{q\perp}\ \ N_q\ ]\Big)$ and its inverse need to be $O_p(1)$. In Example 4, $\widehat{\Psi} = \frac{1}{m-r}(D'_mD_m)^{-1}$ (see Example 2), while Z_T and Z_T^{-1} are easily checked to be bounded in probability. The verification of these conditions is a standard part of computing the asymptotic distributions of cointegration statistics

(Johansen, 1995a,b). See on-line appendices E.3 and E.4 for illustrations of how these conditions are checked in the context of the $I(1)$ model of Example 3 and Example 4 respectively. Al-Sadoon (2016) provides another illustration.

Definition 5 (The Plug–in Principle in General Asymptotics)**.** Suppose $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{P}^{nm}$ are estimators indexed by T, $B^* \in \mathbb{R}^{n\times m}$, and $\widehat{B}^* \in \mathbb{R}^{n\times m}$ is a random sequence indexed by T satisfying $T^{\gamma}(\widehat{B} - \widehat{B}^*) = O_p(1)$. Suppose the null spaces of $\widehat{B}^*$ converge in probability to the null spaces of B^* and let $N_{qT} \in \mathbb{G}^{n\times(n-q)}$ and $M_{qT} \in \mathbb{G}^{m\times(m-q)}$ span the left and right null spaces of $\widehat{B}^*$ respectively. For a given $0 \leq r < \min\{n, m\}$ and RRA, let $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $\widehat{B}_r^{RRA}$ respectively. The weak plug–in principle for statistics of tests of rank is said to hold for the statistic $T^{2\gamma\theta}\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$ relative to the null spaces of $\widehat{B}^*$ if

(i) Under either $H_0(r)$ or $H_T(r)$, $T^{2\gamma\theta}\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) - T^{2\gamma\theta}\tau(\widehat{B}, \widehat{\Omega}, P_{N_{rT}}, P_{M_{rT}}) = O_p(T^{-\gamma})$.

(ii) Under $H_1(r)$, then $|\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})| = O_p^{-1}(1)$ if $|\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r^*}, P_{\widehat{M}_r^*})| = O_p^{-1}(1)$, where $\widehat{N}_r^* \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r^* \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $(\widehat{B}^*)_r^{RRA}$ respectively.

It is said to satisfy the strong plug–in principle relative to the null spaces of $\widehat{B}^*$ if additionally

(iii) Under $H_1(r)$, $\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) - \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r^*}, P_{\widehat{M}_r^*}) = o_p(1)$, where $\widehat{N}_r^* \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r^* \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $(\widehat{B}^*)_r^{RRA}$ respectively. □

When $\gamma = \frac{1}{2}$, $\widehat{B}^*$ is fixed at B^* and N_{qT} and M_{qT} are both fixed and span the null spaces of B^*, the general asymptotics plug–in principle reduces to the standard asymptotics plug–in principle. Note that the correct scaling for our statistics under general asymptotics is $2\gamma\theta$ rather than θ and the quality of the approximation in the general plug–in principle under $H_0(r)$ and $H_T(r)$ depends on the rate at which $\widehat{B}$ converges to the set of matrices of rank r. The factor θ will be determined, just as it was before, by Assumptions K.

The generalized set of assumptions and the generalized notion of the plug–in principle together allow us to generalize Theorem 1.

Theorem 2. Suppose Assumptions K hold along with either Assumptions C or D. Suppose the null space estimators $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ are obtained by either a DBA or a CDA with $\mathrm{cond}(\Theta_T) = O_p(1)$. Let $\widehat{B}^* = [\, N_{q\perp} \quad N_{qT} \,]^{-1\prime} \begin{bmatrix} N_{q\perp}'\widehat{B}M_{q\perp} & 0 \\ 0 & 0 \end{bmatrix} [\, M_{q\perp} \quad M_{qT} \,]^{-1}$ and let $\widehat{N}_r^* \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r^* \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $(\widehat{B}^*)_r^{RRA}$

for $r \leq q$. Suppose (i) the following inclusions hold almost surely

$$P_{\widehat{N}_r^*}\widehat{B}P_{\widehat{M}_r^*} \in \mathcal{X}, \qquad (P_{\widehat{M}_r^*} \otimes P_{\widehat{N}_r^*})\widehat{\Omega}(P_{\widehat{M}_r^*} \otimes P_{\widehat{N}_r^*}) \in \mathcal{Y},$$
$$P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r} \in \mathcal{X}, \qquad (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) \in \mathcal{Y},$$

and (ii) either of the following two conditions hold

(a) For every $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, $N \in \mathbb{G}^{n\times n}$, and $M \in \mathbb{G}^{m\times m}$, $\widetilde{X} = P_{N^{-1}X}N'XMP_{M^{-1}X'} \in \mathcal{X}$, $\widetilde{Y} = (P_{M^{-1}X'}M \otimes P_{N^{-1}X}N)'Y(P_{M^{-1}X'}M \otimes P_{N^{-1}X}N) \in \mathcal{Y}$, and $\kappa(\widetilde{X},\widetilde{Y}) = \kappa(X,Y)$. Moreover, $d(P_{\widehat{N}_{rT}^*}\widehat{B}_TP_{\widehat{M}_{rT}^*},\mathcal{P}) = o_p(1)$ and $d(P_{\widehat{N}_{rT}}\widehat{B}_TP_{\widehat{M}_{rT}},\mathcal{P}) = o_p(1)$, where $\widehat{N}_{rT}^* = [\ N_{q\perp} \quad N_{qT}\ ]^{-1}\widehat{N}_r^*$, $\widehat{M}_{rT}^* = [\ M_{q\perp} \quad M_{qT}\ ]^{-1}\widehat{M}_r^*$, $\widehat{N}_{rT} = [\ N_{q\perp} \quad N_{qT}\ ]^{-1}\widehat{N}_r$, $\widehat{M}_{rT} = [\ M_{q\perp} \quad M_{qT}\ ]^{-1}\widehat{M}_r$, $\widehat{B}_T = [\ N_{q\perp} \quad N_{qT}\ ]'\widehat{B}[\ M_{q\perp} \quad M_{qT}\ ]$.

(b) $Z_T = O_p(1)$, $d(P_{\widehat{N}_r^*}\widehat{B}P_{\widehat{M}_r^*},\mathcal{P}) = o_p(1)$, and $d(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r},\mathcal{P}) = o_p(1)$.

Then $T^{2\gamma\theta}\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}) = T^{2\gamma\theta}\kappa\left(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r},(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\right)$ satisfies the weak plug–in principle relative to the sequence $\widehat{B}^*$. If, additionally, the distance between $\widehat{B}^*$ and the set of discontinuity points of the rank–r RRA is bounded away from zero in probability, then the statistic satisfies the strong plug–in principle relative to $\widehat{B}^*$.

Theorem 2 is strictly more general than Theorem 1. When Assumptions C and D specialize to Assumptions A and B respectively, Theorem 1 is a special case of Theorem 2 (ii). Just as before, condition (i) ensures that the feasible and infeasible statistics are well defined. Condition (ii.a) of Theorem 2 is an invariance condition that allows the plug–in principle to hold in the context of Example 3, where both $\widehat{B}$ and $\widehat{\Omega}$ must be rescaled conformably in order to evaluate the asymptotics. The set of transformations in this condition may seem peculiar. However, they are simple manifestations of the invariance of all of the statistics we have considered so far (except for t) with respect to the group of transformations

$$(\widehat{B},\widehat{\Omega},\widehat{N}_r,\widehat{M}_r) \mapsto (N'\widehat{B}M,(M \otimes N)'\widehat{\Omega}(M \otimes N),N^{-1}\widehat{N}_r,M^{-1}\widehat{M}_r),$$

where $N \in \mathbb{G}^{n\times n}$ and $M \in \mathbb{G}^{m\times m}$. Thus, the set of transformations $(X,Y) \mapsto (\widetilde{X},\widetilde{Y})$ with respect to which κ is invariant in condition (i) defines a group. Condition (ii.b), on the other hand, allows the plug–in principle to hold in the context of Example 4, where Z_T is bounded in probability and the invariance conditions in (i) do not hold. This condition also allows the plug–in principle to hold in standard asymptotics. The intuition of Theorem 2 is the same as in Theorem 1. If the statistic is invariant to rescaling (i.e. condition (ii.a) holds), then the plug–in principle applies to the rescaled statistic exactly as it did before, this time utilizing Lemma 2, because $\kappa\left(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}, (P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\right) = \kappa\left(P_{\widehat{N}_{rT}}\widehat{B}_T P_{\widehat{M}_{rT}}, (P_{\widehat{M}_{rT}}\otimes P_{\widehat{N}_{rT}})\widehat{\Omega}_T(P_{\widehat{M}_{rT}}\otimes P_{\widehat{N}_{rT}})\right)$. If, on the other hand, condition (ii.b) holds, then the same argument as Theorem 1 continues to hold without any modification.

The plug–in principle applied to the context of Example 3 allows one to simply plug–in the limiting null space of $\widehat{B}$. The plug–in principle applies regardless of the order of integration of the process (fractional cointegration is also allowed) and not only to the F statistic but also to the J_g and LR_g statistics. It also applies in the contexts of added lags and arbitrary deterministic terms such as polynomial trends and dummies. The plug–in principle in Example 4 applies relative to a random sequence rather than a constant one. In particular, one cannot plug–in the null space of B^*. That is because under either $H_0(r)$ or $H_T(r)$, $P_{\widehat{M}_r} - P_{M_r} = O_p(T^{-1/2})$, which is too slow for the plug–in principle to work. One can, however, plug in M_{rT} as defined in Example 4 because $P_{\widehat{M}_r} - P_{M_{rT}} = O_p(T^{-1})$. See on-line appendices E.4 and F.4 for theoretical and Monte Carlo illustrations respectively.

A large class of statistics is nested under Theorem 2, including all of the standard asymptotics statistics of the literature as well as the majority of the cointegration rank statistics in the literature. In particular, it nests all of the statistics included in Table 1 except for the ones superscripted by the symbol †. Those, along with recent statistics by Hallin et al. (2016) and Boswijk et al. (2015), are of the form $T^{2\gamma\theta}\tau(\{y_t : t = 1, \ldots, T\}, P_{\widehat{M}_r})$. Thus, they explicitly depend on a null space estimator and their dependence on the data is more complicated than what we have considered in this paper. However, it is evident from the proofs of the asymptotics of these results that these statistics are asymptotically equivalent to infeasible versions $T^{2\gamma\theta}\tau(\{y_t : t = 1, \ldots, T\}, P_{M_r})$ under $H_0(r)$ and that both the feasible and infeasible statistics diverge under $H_1(r)$. Thus, a form of the plug–in principle continues to hold for these statistics as well.

It follows from Theorem 2 that the Johansen (1988), Kleibergen & van Dijk (1994), and Kleibergen & Paap (2006) statistics, which differ from each other only in their implicit null space estimators, differ from each other by $O_p(T^{-1/2})$ under $H_0(r)$ and $H_T(r)$. Thus, the choice among these will have to depend on either Monte Carlo performance or numerical expedience as noted in Section 3.

The simplification to asymptotic analysis afforded by Theorem 2 is noteworthy. It allows the researcher to obtain the asymptotics not only for the different alternatives but also under misspecification. We summarize in the following corollaries.

Corollary 3. Suppose Assumptions K and C hold and suppose we have null space estimators $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ obtained by either a DBA or a CDA with $\operatorname{cond}(\Theta_T) = O_p(1)$. Under $H_0(r)$ or $H_T(r)$, if $T^{2\gamma\theta}\tau\left(\widehat{B}, \widehat{\Omega}, P_{N_{rT}}, P_{M_{rT}}\right) \xrightarrow{d} \zeta$, then $T^{2\gamma\theta}\tau\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \zeta$. In particular, if

$$\left(T^{\gamma}\operatorname{vec}(N_{rT}'\widehat{B}M_{rT}), (M_{rT}\otimes N_{rT})'\widehat{\Omega}(M_{rT}\otimes N_{rT})\right) \xrightarrow{d} (\xi_r, \Omega_r),$$

then we have

$$T^{2\gamma-1}LR_g\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\xi_r\|^2_{\Omega_r}, \qquad T^{2\gamma-1}F\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\xi_r\|^2_{\Omega_r},$$
$$T^{2\gamma-1}J_g\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\operatorname{mat}(\Omega_r^{-1/2}\xi_r)\|_2^2.$$

Under $H_0(r)$, ξ_r and Ω_r in Corollary 3 are typically functionals of a Brownian motion and deterministic terms (if deterministic trends are included), while under $H_T(r)$ they are typically of the Ornstein–Uhlenbeck form (Hubrich et al., 2001). The limiting behaviour under $H_0(r)$ and $H_T(r)$ of all of the statistics in Johansen (1988), Johansen (1991), Kleibergen & van Dijk (1994), Yang & Bewley (1996), Quintos (1998), Gonzalo & Pitarakis (1999), Lütkepohl & Saikkonen (1999), Kleibergen & Paap (2006), Avarucci & Velasco (2009), and Cavaliere et al. (2010a) follow from Corollary 3. These results assume correct specification, so the limiting distributions above are nuisance–parameter–free. In the case of misspecification, the limiting distributions may not be free of nuisance parameters. It follows from Corollary 3 that the statistics proposed by Johansen (1988), Kleibergen & van Dijk (1994), and Kleibergen & Paap (2006) have the exact same behaviour under the misspecification conditions of Caner (1998) (infinite variance shocks), Cavaliere et al. (2010b) (heteroskedastic shocks), and Aznar & Salvador (2002) and Cavaliere et al. (2014) (misspecified lag length).

Corollary 4. Suppose Assumptions K and D hold and we have a null space estimator $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ obtained by either a DBA or a CDA with $\operatorname{cond}(\Theta_T) = O_p(1)$. Under $H_0(r)$ or $H_T(r)$, if $T^{2\gamma\theta}\tau\left(\widehat{B}, \widehat{\Omega}, P_{M_{rT}}, P_{M_{rT}}\right) \xrightarrow{d} \zeta$, then $T^{2\gamma\theta}\tau\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \zeta$. In particular, if

$$\left(T^{\gamma}\operatorname{vech}(M_{rT}'\widehat{B}M_{rT}), D_{m-r}^{\dagger}(M_{rT}\otimes M_{rT})'\widehat{\Omega}(M_{rT}\otimes M_{rT})D_{m-r}^{\dagger}{}'\right) \xrightarrow{d} (\xi_r, \Omega_r),$$

then we have

$$T^{2\gamma-1}LR_g\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\xi_r\|^2_{\Omega_r}, \qquad T^{2\gamma-1}F\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\xi_r\|^2_{\Omega_r},$$
$$T^{2\gamma-1}J_g\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \|\operatorname{mat}(D_{m-r}\Omega_r^{-1/2}\xi_r)\|_2^2,$$

and if $M_{rT} \xrightarrow{p} M_r$, a non–random matrix of orthonormal columns, then

$$T^{\gamma-1/2}t\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}\right) \xrightarrow{d} \frac{\operatorname{tr}(\operatorname{mat}(D_{m-r}\xi_r))}{(\operatorname{vec}'(I_{m-r})D_{m-r}\Omega_r D_{m-r}'\operatorname{vec}(I_{m-r}))^{1/2}}.$$

Similar observations apply to Corollary 4 as do to Corollary 3. The asymptotic distributions of the Bierens (1997), Nyblom & Harvey (2000), Breitung (2002), and Nielsen (2010) statistics under $H_0(r)$ and $H_T(r)$ follow from Corollary 4.

Corollaries 3 and 4 (and by inclusion, Corollaries 1 and 2) highlight a remarkable property of the asymptotics of statistics for tests of rank that does not seem to have received sufficient attention: only the asymptotics along N_{rT} and M_{rT} have any contribution to the limiting distribution of the test statistics. That is, the asymptotics of $\widehat{B}$ along neither $N_{r\perp}$ nor $M_{r\perp}$ have any contribution to the asymptotics whatsoever.

5 Monte Carlo

This section illustrates the plug–in principle with two simulation experiments based on Examples 1 and 4. These were chosen in order to introduce fixed–b tests in the first example and to showcase the tests based on the QR and Cholesky RRAs in the second example. Size and power tables along with other simulations based on Example 2 and 3 are available in on-line appendix E.

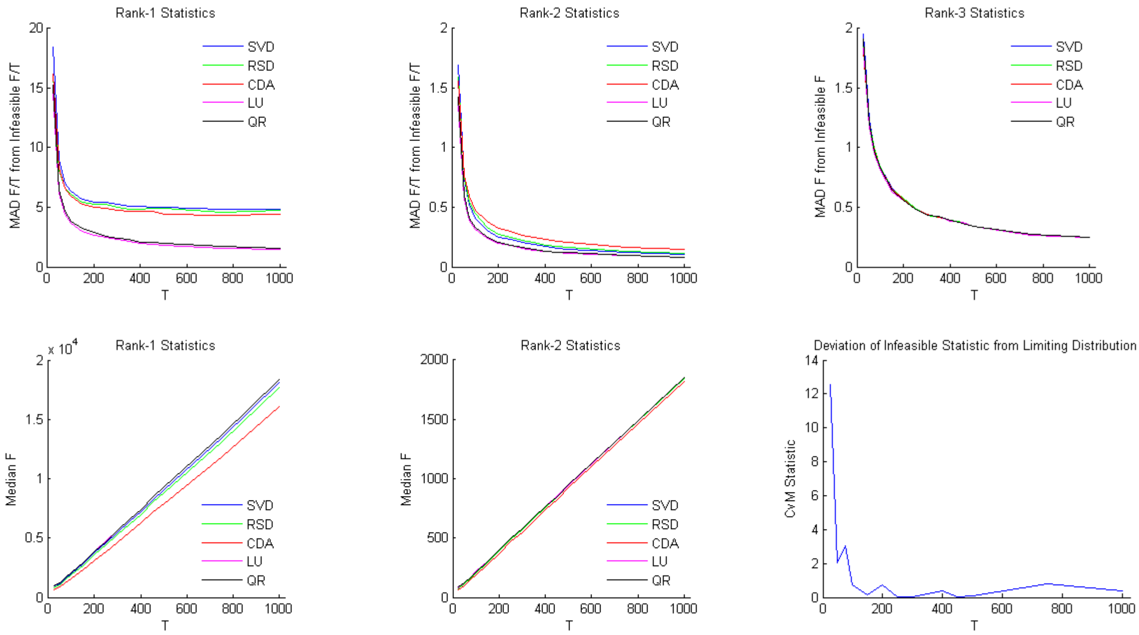
5.1 Linear Regression

Considering the set up of Example 1. Let $\{x_t : t = 1, \ldots, T\}$ be i.i.d. $N(0, I_4)$ and independent of $\{\varepsilon_t : t = 1, \ldots, T\}$, a stationary 4–dimensional process satisfying $\varepsilon_t = 0.5\varepsilon_{t-1} + u_t$. Let $B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. For $\widehat{\Omega}$, we use the non–parametric estimator based on the Bartlett kernel with bandwidth equal to the sample size. The experiment consists of generating 10000 samples of sizes ranging from $T = 25$ to $T = 1000$. For each sample, we compute $F(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$ for $r = 0, \ldots, 3$ and null spaces estimated by SVD, RSD, CDA, LU, and QR RRAs. In the case of the CDA, we set $\Theta = \widehat{\Omega}$. We also consider the infeasible F statistics based on the SVD.

The right panels of Figure 1 illustrate the plug–in principle under $H_0(3)$. The top right panel shows the median absolute difference between each feasible statistic and the infeasible statistic goes to zero in the limit. The lower right panel then provides the Carmér–von Mises distance between the distribution of the infeasible statistic and the asymptotic distribution, $\frac{W^2(1)}{2\int_0^1 (W(s)-sW(1))^2 ds}$ for a standard Brownian motion W (Kiefer & Vogelsang, 2002a). Thus, all our rank–3 statistics also have this limiting distribution.

The middle panels illustrates the strong plug–in principle under $H_1(2)$. In particular, the top middle panel plots the median absolute deviation of each feasible statistic from the

Figure 1: Monte Carlo Results for F Statistics



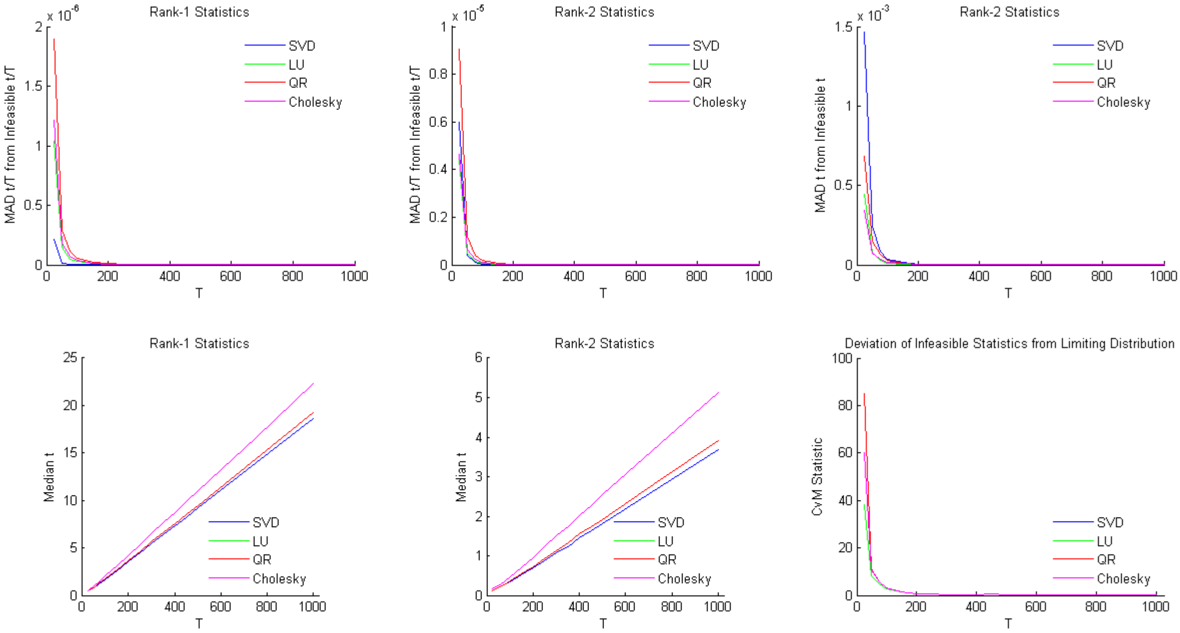
associated infeasible statistic divided by T. The rates of divergence of the statistics are asymptoticallly equal. This is clearly visible in the lower middle panel which plots the median F statistic for each sample size. This is due to the fact that $B_2^{RRA} = \begin{bmatrix} 1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0 \end{bmatrix}$ for all of our RRAs and all of our rank–2 RRAs are continuous at B.

Finally, the left panels illustrate a setting where the weak plug–in principle holds but not the strong plug–in principle. The top right panel plots the rates of divergence which are now heterogeneous. This is also visible in the bottom left panel which plots the median statistics. This is due to the fact that none of the rank–1 RRAs are continuous at B. $\widehat{B}_1^{RRA}$ fluctuates between the vicinity of $B_1^{RRA} = \begin{bmatrix} 1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0 \end{bmatrix}$ and the vicinity of $B_2^{RRA} = \begin{bmatrix} 0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0 \end{bmatrix}$ as $\widehat{B} \xrightarrow{p} B$ for each RRA. Thus, the null space estimators fluctuate as well and F/T fails to converge although it stays bounded away from zero in probability by the weak plug–in principle.

5.2 The Local Level Model

Consider the model given in Example 4 with $\Sigma = I_4$, $B = \begin{bmatrix} 1&0&0&0\\0&1&0&0\\0&0&0.5&0\\0&0&0&0 \end{bmatrix}$. As before, the experiment consists of generating 10000 samples of sizes ranging from $T = 25$ to $T = 1000$. For each sample, we compute the non–standardized t statistic $\sqrt{T}t\left(\widehat{B}, \frac{1}{m-r}(D_m'D_m)^{-1}, P_{\widehat{M}_r}\right) = T\mathrm{tr}(P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r})$, for $r = 0, \ldots, 3$. In particular, we compute the statistics based on the SVD, LU, QR, and Cholesky RRAs along with their infeasible analogues. The Cholesky statistics appear here for the first time. Note that the SVD statistic is the Nyblom & Harvey (2000)

Figure 2: Monte Carlo Results for Cointegration t Statistics



statistic. Note also that the infeasible statistics do not agree across RRAs because, for $r = 1, 2$, $(\widehat{B}^*)_r^{RRA}$ is not invariant to the choice of RRA as it was in the previous experiment.

The format of Figure 2 is the same as in 1. In particular, the right panels of Figure 2 show that the difference between each statistic and its infeasible analogue goes to zero in the limit. In turn, all of our statistics have the same limiting distribution under $H_0(3)$ (see on-line appendix E). The middle and left panels then look at the performance under $H_1(2)$ and $H_1(1)$ respectively. The top panels make clear that the strong plug–in principle is at play here as the median absolute difference between each statistic and its infeasible analogue goes to zero. This is due to the fact that each vech$(\widehat{B}^*)$ is non–degenerate in the space of positive semi–definite matrices of rank 3. Thus $(\widehat{B}^*)_r^{RRA}$ stays away from the (non–generic) points of discontinuity of each RRA for $r = 1, 2$. However, unlike in the previous experiment, the rates of divergence displayed in the lower panels are heterogeneous, again, because $(\widehat{B}^*)_r^{RRA}$ differ across RRAs for $r = 1, 2$.

5.3 Further Monte Carlo

More extensive experiments are provided in on-line appendix F. These consider size and power of the various tests above in the context of Examples 1 – 4. Here we summarize the findings. First, there does not appear to be a uniformly best performing test in small samples. Second, tests based on the infeasible statistics have a tendency to over–reject relative to the tests based on the feasible statistics. To see why this makes sense, recall that statistics for tests of rank can be seen as measures of how large $P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}$ is (see Section 3.1). When this quantity is small, we accept $H_0(r)$ and when it is large we reject in favour of $H_1(r)$. Now under $H_0(r)$, $P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}$ is likely to be of smaller size than $P_{N_r}\widehat{B}P_{M_r}$ because in the former expression the null space estimators are actively trying to annihilate $\widehat{B}$ whereas is the latter expression the implicit null space estimators are passive and do not adapt to $\widehat{B}$. Third, fixed–b tests have better size properties than their small–b counterparts. This accords with the Monte Carlo evidence documented in the fixed–b literature. Finally, the Cragg & Donald (1997) tests have a tendency to reject less frequently than all the other statistics regardless of the alternative being tested.

6 Conclusion

This paper has demonstrated that the asymptotic behaviour of statistics for tests of rank is determined by the asymptotic behaviour of implicit null space estimators through a plug–in principle. This has allowed for a general theory of tests of rank that simplified the asymptotics under the various alternatives, clarified the relationships between the various statistics in the literature, made full use of the numerical analysis literature, and motivated many new tests.

We briefly mention some possible venues for future research. First, we have considered the performance of test of rank under a fixed–rank hypothesis and higher rank alternatives but not lower rank alternatives. The approach of this paper can be used to obtain well known results by Cragg & Donald (1997) (Al-Sadoon, 2015, p. 40) but generalizations of these results seem to be highly non–trivial and deserve further investigation. Second, as a number of statistics in the literature have been shown to be asymptotically equivalent, the next natural step is to study small sample performance and higher order asymptotics. On-line appendix F goes a small step in that first direction, while the second direction is the subject of ongoing research. Third, as this paper has presented a theory of null space estimation, the natural next step is to consider inference on the directionality (rather than the dimension) of the null spaces. Fourth, model selection techniques can be used to turn any statistic for a test of rank into an index whose optimum estimates rank (see Al-Sadoon (2015)), thus rank estimation deserves further investigation in both theory and Monte Carlo simulations. Finally, high–dimensional tests of rank would be very useful for many applications in macroeconomics and finance.

References

Al-Sadoon, M. M. (2015). A General Theory of Rank Testing. Working Papers 750, Barcelona Graduate School of Economics.

Al-Sadoon, M. M. (2016). Testing Subspace Granger Causality. Working Papers 850, Barcelona Graduate School of Economics.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, *22*, 327–351.

Andrews, D. W. K. (1987). Asymptotic results for generalized wald tests. *Econometric Theory*, *3*(3), 348–358.

Avarucci, M. & Velasco, C. (2009). A wald test for the cointegration rank in nonstationary fractional systems. *Journal of Econometrics*, *151*(2), 178–189.

Aznar, A. & Salvador, M. (2002). Selecting the rank of the cointegration space and the form of the intercept using an information criterion. *Econometric Theory*, *18*(4), pp. 926–947.

Bartlett, M. S. (1947). Multivariate analysis. *Supplement to the Journal of the Royal Statistical Society*, *9*(2), 176–197.

Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics*, *77*(2), 379–404.

Boswijk, H. & Lucas, A. (2002). Semi-nonparametric cointegration testing. *Journal of Econometrics*, *108*(2), 253 – 280.

Boswijk, H. P., Jansson, M., & Nielsen, M. O. (2015). Improved likelihood ratio tests for cointegration rank in the var model. *Journal of Econometrics*, *184*(1), 97 – 110.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of Econometrics*, *108*(2), 343–363.

Camba-Mendez, G. & Kapetanios, G. (2009). Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. *Econometric Reviews*, *28*(6), 581–611.

Caner, M. (1998). Tests for cointegration with infinite variance errors. *Journal of Econometrics*, *86*(1), 155 – 175.

Cavaliere, G., De Angelis, L., Rahbek, A., & Taylor, A. R. (2014). Determining the cointegration rank in heteroskedastic var models of unknown order. Mimeo.

Cavaliere, G., Rahbek, A., & Taylor, A. R. (2010a). Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory*, *26*(06), 1719–1760.

Cavaliere, G., Rahbek, A., & Taylor, A. R. (2010b). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics*, *158*(1), 7–24.

Cragg, J. G. & Donald, S. G. (1996). On the asymptotic properties of ldu-based tests of the rank of a matrix. *Journal of the American Statistical Association*, *91*(435), 1301–1309.

Cragg, J. G. & Donald, S. G. (1997). Inferring the rank of a matrix. *Journal of Econometrics*, *76*(1-2), 223–250.

Donald, S. G., Fortuna, N., & Pipiras, V. (2007). On rank estimation in symmetric matrices: The case of indefinite matrix estimators. *Econometric Theory*, *23*(06), 1217–1232.

Donald, S. G., Fortuna, N., & Pipiras, V. (2014). On estimating the rank of semide
nite matrix. Mimeo.

Dufour, J.-M. & Valery, P. (2011). Hypothesis tests when rank conditions fail: a smooth regularization approach. mimeo, HEC Montréal.

Duplinskiy, A. (2014). Is regularization necessary? a wald-type test under non-regular conditions. Research Memorandum 025, Maastricht University, Graduate School of Business and Economics.

Gabriel, K. R. & Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, *21*, 489–498.

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations* (3 ed.). Classics in Applied Mathematics. Baltimore, USA: Johns Hopkins University Press.

Gonzalo, J. & Pitarakis, J.-Y. (1999). Dimensionality effect in cointegration analysis. In R. Engle & H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger* (pp. 212–229). Oxford University Press.

Hallin, M., van den Akker, R., & Werker, B. J. (2016). Semiparametric error-correction models for cointegration with trends: Pseudo-gaussian and optimal rank-based tests of the cointegration rank. *Journal of Econometrics*, *190*(1), 46 – 61.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hansen, P. C. (1998). *Rank-Deficient and Discrete Ill-Posed Problems*. Monographs on Mathematical Modelling and Computation. Philadelphia, USA: Society for Industrial and Applied Mathematics.

Hubrich, K., Lütkepohl, H., & Saikkonen, P. (2001). A Review Of Systems Cointegration Tests. *Econometric Reviews*, *20*(3), 247–318.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2-3), 231 – 254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, *59*(6), 1551–1580.

Johansen, S. (1995a). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford, United Kingdom: Oxford University Press.

Johansen, S. (1995b). A stastistical analysis of cointegration for I(2) variables. *Econometric Theory*, *11*(01), 25–59.

Johansen, S. & Nielsen, M. O. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, *80*(6), 2667–2732.

Kiefer, N. M. & Vogelsang, T. J. (2002a). Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica*, *70*(5), 2093–2095.

Kiefer, N. M. & Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, *18*(06), 1350–1366.

Kiefer, N. M. & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, *21*(06), 1130–1164.

Kiefer, N. M., Vogelsang, T. J., & Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, *68*(3), 695–714.

Kleibergen, F. & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126.

Kleibergen, F. & van Dijk, H. K. (1994). Direct cointegration testing in error correction models. *Journal of Econometrics*, *63*(1), 61–103.

Lehmann, E. L. & Romano, J. P. (2005). *Testing Statistical Hypotheses* (3 ed.). Springer Texts in Statistics. New York, USA: Springer.

Lucas, A. (1997). Cointegration Testing Using Pseudolikelihood Ratio Tests. *Econometric Theory*, *13*(02), 149–169.

Lucas, A. (1998). Inference on cointegrating ranks using lr and lm tests based on pseudo-likelihoods. *Econometric Reviews*, *17*(2), 185–214.

Lütkepohl, H. & Burda, M. M. (1997). Modified wald tests under nonregular conditions. *Journal of Econometrics*, *78*(2), 315–332.

Lütkepohl, H. & Saikkonen, P. (1999). A lag augmentation test for the cointegrating rank of a var process. *Economics Letters*, *63*(1), 23 – 27.

Markovsky, I. (2012). *Low Rank Approximation: Algorithms, Implementation, Applications*. Communications and Control Engineering. New York, USA: Springer.

Massmann, M. (2007). Reduced rank regressions and the poincaré separation theorem. Mimeo, Institute of Econometrics. University of Bonn.

Moore, D. S. (1977). Generalized inverses, wald's method, and the construction of chi-squared tests of fit. *Journal of the American Statistical Association*, *72*(357), 131–137.

Nielsen, M. O. (2010). Nonparametric cointegration analysis of fractional systems with unknown integration orders. *Journal of Econometrics*, *155*(2), 170–187.

Nyblom, J. & Harvey, A. (2000). Tests of common stochastic trends. *Econometric Theory*, *16*(02), 176–199.

Quintos, C. E. (1998). Fully modified vector autoregressive inference in partially nonstationary models. *Journal of the American Statistical Association*, *93*(442), pp. 783–795.

Reinsel, G. C. & Velu, R. P. (1998). *Multivariate Reduced-Rank Regression*. Lecture Notes in Statistics. New York, USA: Springer.

Robin, J. & Smith, R. (1995). Tests of rank. Cambridge Working Papers in Economics 9521, Faculty of Economics, University of Cambridge.

Robin, J.-M. & Smith, R. J. (2000). Tests of rank. *Econometric Theory*, *16*(02), 151–175.

Stewart, G. W. & Sun, J. (1990). *Matrix Perturbation Theory*. New York, USA: Academic Press, Inc.

Stock, J. H. & Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, *83*(404), pp. 1097–1107.

White, H. (1994). *Estimation, Inference, and Specification Analysis*. Econometric Society Monographs. Cambridge, UK: Cambridge University Press.

Wright, J. H. (2003). Detecting lack of identification in gmm. *Econometric Theory*, *19*(02), 322–330.

Yang, M. & Bewley, R. (1996). On cointegration tests for var models with drift. *Economics Letters*, *51*(1), 45 – 50.